In-silico design novel phenylsulfonyl furoxan and phenstatin derivatives as multi-target anti-cancer inhibitors based on 2D-QSAR, molecular docking, dynamics and ADMET approaches

Abdelmadjid Guendouzi, Lotfi Belkhiri, Farah Djelti, Zineddine Mohamed Zendaoui, Houari Brahim, Abdelkrim Guendouzi, Abdelhamid Djekoun, Abdou Boucekkine

In silico design novel Phenylsulfonyl Furoxan and Phenstatin derivatives as multi-target anti-cancer inhibitors based on 2D-QSAR, molecular docking, dynamics and ADMET approaches.

Abdelmadjid Guendouzi,[a,b] Lotfi Belkhiri,[b,c*] Farah Djelti,[d] Zineddine Mohamed Zendaoui,[b] Houari Brahim,[e] Abdelkrim Guendouzi,[e] Abdelhamid Djekoune,[b], Abdou Boucekkine[f]

[a] *Ecole Normale Supérieure ENS Assia Djebar Constantine, Ali Mendjeli, 25000 Constantine, Algeria*
[b] *Centre de Recherche en Sciences Pharmaceutiques CRSP, ZAM Ali Mendjeli, Constantine, Algeria*
[c] *Laboratoire de Physique Mathématique et Subatomique LPMS, Département de Chimie, Université des Frères Mentouri, 25017 Constantine, Algeria.*
[d] *Département de biologie, Faculté des sciences de la nature et de la vie et sciences de la terre et de l'univers, Université de Tlemcen 13000, Algeria*
[e] *Laboratoire de Chimie, Synthèse, Propriétés et Applications LCSPA, Faculté des Sciences, Université de Saida, Algeria*
[f] *Univ Rennes, ISCR, UMR 6226 CNRS, Campus de Beaulieu, 35042 Rennes Cedex, France*

Abstract

A series of thirty-one hybrid of phenylsulfonyl furoxan and phenstatin (1a-j, 2a-j, 3a-j, 4, and 5) derivatives, were computationally studied as potential anti-cancer inhibitors against four cell lines, i.e., A2780, MDA-MB-231, HCT-116, and A549. In this work, the 2D-QSAR approach combining the multiple linear regression (MLR) model, and internal and external cross-validation, showed a satisfactory quality factor: R^2 = 0.85, 0.74, 0.82, and 0.75 for the four cell lines, respectively. The binding affinity of the hybrid agents towards the four 4GL7, 6GUE, 1M17, and 4XL7 antitumoral targets, was further evaluated using molecular docking and dynamics simulations (0 - 200 ns). The dynamics assessment parameters indicated the formation of satisfactorily stable complexes. In addition, all considered data sets show that the best binding affinity, including the highest docking score, hydrogen bond energy, and amino acid steric interactions, are well predicted for the best-selected complexes. The developed 2D-QSAR model was leveraged to design and predict the biological activity of twelve new hybrid compounds (N1–N12) based on the best in vivo inhibitor, namely, the 3h ligand of the formula: (4-((1-(2-((4-((3crylamidophenyl)amino)quinazolin-2-yl)thio)acetyl)piperidin-4-yl)oxy)-3-(phenylsulfonyl)-1,2,5-oxadiazole 2-oxide). Multitargeting docking scores and dynamics simulations show that they exhibit satisfactorily potent antitumoral inhibition abilities towards the four proteins. Our *in-silico* outcomes would be combined with in-vitro and in-vivo studies to provide a perspective on the validation of their anticancer activity. In particular, the ADMET predictions indicate that four new designed ligands have demonstrated a good drug-like profile, and can be considered prospective candidates for future anti-cancer therapies.

Keywords: Furoxan derivatives, Multi-target, anti-cancer, QSAR, Molecular docking, Dynamics, ADMET.

Corresponding Authors: *E-mail (Lotfi Belkhiri): lotfi.belkhiri@umc.edu.dz

1. Introduction

Nowadays, cancer remains a significant public health challenge in the world, with millions of individuals being diagnosed with various forms of the disease each year. Ovarian, colon, prostate, and lung adenocarcinoma diseases are among the most common and deadly types of cancer. According to the recent American Cancer Society (ACS) [1] and the World Health Organization (WHO) report [2], more than half a million persons are expected to die from cancer in 2023 [3].

Although numerous progress in research, prevention efforts, and treatment options that have led to improved outcomes for patients, such cancers are still challenging, both at the cost level and in terms of drugs availability [4]. Over the two past decades, to reduce the mortality rates of cancer, research into the mechanisms and the development of new diagnostic tools, based on multitargeting approaches for the design of novel anticancer drugs, have gained great interest [5].

Furthermore, the cost of drug development has sharply increased along with the high rate of clinical trial failures. Such an increase in expenses is partially due to the inability of the "one drug – one target" approach to predict drug side effects and toxicity [6].

Therefore, multitargeted treatments represent a new paradigm in the field of cancer therapy. In contrast to traditional chemotherapy, these therapeutic approaches exhibit a higher degree of specificity in targeting cells that are directly associated with tumors proliferation.

To tackle this issue, polypharmacology as an alternative approach, confirms that multi-target drugs are more efficient ways of treating them as opposed to single-target drugs [7]. Currently, polypharmacology aims to study small molecule interactions with multiple targets. However, apart from developing more potent and effective multitargeting drugs, exhaustive polypharmacology in-vitro or in-vivo studies are not easily practical, and to design multi-target ligands, several factors and challenges must be taken into account [6].

Faced with tremendous challenges, wide efforts have led to a variety of novel approaches for predicting drug design polypharmacology. Indeed, in recent years, computational or in silico polypharmacology has gained significant attention to study the promiscuous nature of drugs. Moreover, in silico methods can substantially widen this search by providing tools able to predict the best drug-target binding affinities and rationalize the process by reducing duration, cost, and attrition rate [8] and have been recently approved by the FDA (Food & Drug Agency) organization [9].

In particular, computational polypharmacology has gained significant attention in studying the promiscuous nature of drugs [10], and recent achievements in the computational design of

multitargeting drugs have established this computational approach as a promising alternative approach to predicting unknown targets or side effects [11].

In the present study, a series of thirty-one hybrids of phenylsulfonyl furoxan and phenstatin derivatives (**1a-j**, **2a-j**, **3a-j**, **4** and **5**), based on their observed biological (IC50/µM) [12], are computationally investigated for the first time for their potent anti-cancer inhibition candidates against breast, colorectal, lung, and ovarian diseases. As reported by the same authors, in-vitro evaluation of their anti-tumor activity shows that most of such hybrid derivatives feature significantly enhanced anti-proliferation activities against several human cancer cell lines, i.e., A2780, MDA-MB-231, HCT-116, and A549. Importantly, among the synthesized molecules, compound **3h** with formula {4-(2-(5-(4-bromobenzoyl)-2-ethoxyphenoxy)ethoxy)-3-(phenylsulfonyl)-1,2,5-oxadiazole 2-oxide}, as shown on figure 1, exhibited the most potent anti-tumor activities against several human cancer cell lines, with IC50 values ranging from 0.008 to 0.021 µM [12].

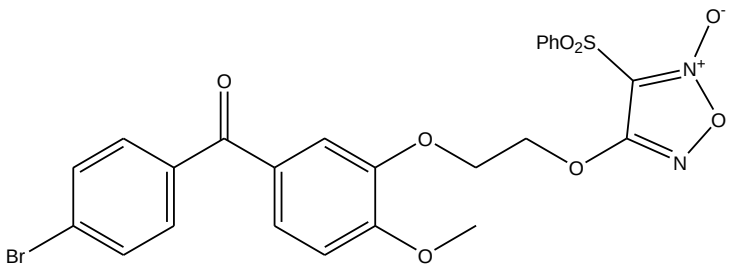


Figure 1: Molecular structure of **3h** compound

Referring to the best inhibition in-vivo score of such **3h** compounds, in silico approaches, including 2D-QSAR, and molecular docking in combination with dynamic simulation techniques, were used for the four cancer cell lines, to rationalize the ability of thirty-one ligands towards the inhibition of multiple targets involved in anti-tumor development, and allows to predict novel potent multitargeting inhibition candidates.

Statistical techniques, such as multilinear regression (MLR) [13], internal and external validation such as Y-randomization methods, and applicability domain (AD) [14, 15], were used to validate the 2D-QSAR model. Additionally, molecular docking in combination with dynamics simulations were carried out to predict the best ligand-protein binding affinity and stability of the ligand-receptor complexes.

The results obtained from the in silico study aimed to provide valuable insights into the discovery and design of new twelve (N1– N12) inhibitors for the treatment of cancer, and

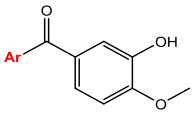
their perspective development via in-vitro and in-vivo assays, particularly those related to multitargeting drugs approaches.

2. Materials and methods

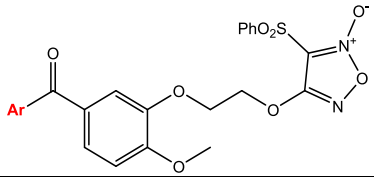
2.1 QSAR modeling

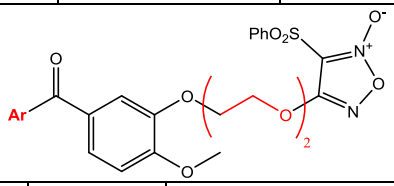
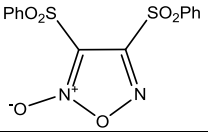
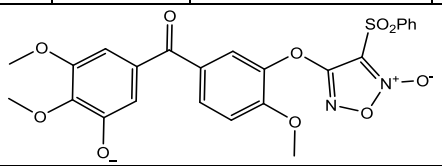
To evaluate computationally the anticancer activity of hybrid phenylsulfonyl furoxan and phenstatin derivatives against the four different human tumor cell lines, i.e., ovarian (A2780), breast (MDA-MB 231), lung adenocarcinoma (A549), and colorectal (HCT-116) [12], 2D-QSAR approach was first used for the actual thirty-one (**1a-j**, **2a-j**, **3a-j**, **4** and **5**) compounds with substituted (Ar) chemical structures and their observed biological activity reported in Table 1. The structures were generated using Avogadro software [16], and then optimized using Density Functional Theory (DFT) calculations with the AMS2022.107 release software [17] using the B3LYP functional [18, 19] coupled to the polarized triple zeta TZP basis set as implemented in the AMS/ADF engine. To ensure that the optimized structures were true minima on the potential energy surface, vibrational frequency calculations were performed at the same B3LYP/TZP level of theory, no imaginary frequency being observed.

Table 1: Chemical structures of phenylsulfonyl furoxan and phenstatin derivatives with their experimental biological activities (pIC50) related to the four cell lines A2780, MDA-MB-231, HCT 116, and A549[12]. (*) test set.



Ligand (pIC50)	A2780	MDA -MB-231	HCT-116	A549	Ar
1a*	6.92	6.29	5.77	5.77	3, 4, 5-trimethoxyphenyl
1b	5.35	4.91	4.25	4.49	2-fluorophenyl
1c	5.53	4.33	4.45	4.28	3-fluorophenyl
1e	5.43	4.38	4.58	5.77	4-chlorophenyl
1f	5.49	4.49	4.32	4.49	2-bromophenyl
1g	6.34	5.25	4.9	4.28	3-bromophenyl
1h*	5.94	4.68	4.44	5.77	4-bromophenyl
1i	6.77	5.92	4.91	4.49	2-naphthalenyl
1j	4.91	3.88	3.84	4.28	6-bromonaphthalen-2-yl



2a	7.46	6.08	5.99	5.99	3, 4, 5-trimethoxyphenyl
2b*	6.28	5.14	4.44	4.44	2-fluorophenyl
2c	6.54	5.77	5.62	5.62	3-fluorophenyl
2d	6.82	5.68	5.8	5.8	
2e	7.39	5.96	6.08	6.08	4-chlorophenyl
2f	6.82	5.27	5.54	5.54	2-bromophenyl
2g	7.39	5.82	5.37	5.37	3-bromophenyl
2h*	7.92	6.82	6.96	6.96	4-bromophenyl
2i*	6.92	5.44	6.02	6.02	2-naphthalenyl
2j	7.54	6.25	5.96	5.96	6-bromonaphthalen-2-yl
					
3a	6.77	6.14	5.49	5.49	3, 4, 5-trimethoxyphenyl
3b	6.49	5.32	5.29	5.29	2-fluorophenyl
3c*	6.28	5.14	4.44	4.44	3-fluorophenyl
3d	6.46	5.21	5.04	5.04	
3e	6.62	4.92	4.77	4.77	4-chlorophenyl
3f	6.26	4.95	4.66	4.66	2-bromophenyl
3g	6.47	5.23	4.92	4.92	3-bromophenyl
3h	6.82	5.18	5.33	5.33	4-bromophenyl
3i	6.36	5.42	4.64	4.64	2-naphthalenyl
3j*	6.59	5.57	5.21	5.21	6-bromonaphthalen-2-yl
					
4*	6.06	4.73	4.57	4.57	-
					
5	6.04	4.94	4.64	4.64	-

To build the 2D-QSAR model, the thirty-one selected series of compounds were divided into a training set (23 compounds) and a test set (8 compounds). All activity values (IC50, μM) for the multi-target compounds were converted into negative logarithms (pIC50 = 6 – logIC50) and used as dependent variables for the 2D-QSAR models. The molecular structures of the thirty-one selected compounds and their converted pIC50 are reported in the Table 1.

2.2. Descriptor generation and MLR analysis

For each structure, over 2000 descriptors were calculated using Dragon software [20], including unidimensional (1D) descriptors as well as 2D and 3D ones. These latter encode a wide range of molecular properties, such as physicochemical, electronic, and topological properties. Some descriptors that might be calculated include molecular weight, boiling point, polarizability, electronegativity, molecular volume, and many others. Multiple linear regression (MLR) techniques provide useful statistical tools that quantify the relationship between dependent and independent variables.

The R [21] was used to obtain the 2D-QSAR models, using the multilinear regression fits with linear model of the form:

$$Y = \beta_0 + \sum_{i=1} \beta_i x_i + e \qquad (1)$$

Where Y and x_i are the dependent (pIC50) and independent (molecular descriptors) variables, respectively, and the parameters: β_0 and β_i are the Regression coefficients.

The internal validation method using cross-validation leave-one-out Q^2 parameter (LOOCV) [22] combined with several performance metrics, such as the coefficient of determination R^2, the adjusted coefficient of determination R^2_{adj}, the mean square error of the model MSE, Y-randomization parameters R^2_{rand} [14], the coefficient of determination of the test set R^2_{test}, and applicability domain (AD), were used to validate the obtained 2D-QSAR model and to ensure its robustness and its reliability [23, 24].

Importantly, Y-randomization tests are a statistical method used to evaluate the validity of the 2D-QSAR models, so that the biological activity values (pIC50) are randomly shuffled or permuted while keeping the chemical structure of the compounds unchanged. Then, the 2D-QSAR model is trained using the permuted activity values, and the resulting model is used to predict the original (not permitted) activity values. The process is repeated multiple times, and the resulting R^2_{rand} values are compared to the original R^2 value obtained from the model trained on the not permutated data.

The applicability domain approach helps to identify outliers or influential molecules and defines the range of compounds for which the 2D-QSAR model is valid [25]. To define and describe the applicability domain of the built 2D-QSAR model, the leverage h_i method [14] is used and is given by Equation 2:

$$hi = X_i(X^TX)\text{-}1X_iT \qquad (2)$$

Where i is the training molecule set, X is the nk descriptor matrix of the training set molecule, and X^T is the transpose of the training set (X). X_iT is the transpose matrix X_i used to build the model.

The critical leverage h* is defined as follows:

$$h^* = 3(j + 1)/m \quad (3)$$

Where j is the number of descriptors in the built model, and m is the number of training compounds set. The critical leverage is the limit value to check the best 2D-QSAR model that will be used to design new molecules with good predicted anticancer activity values [26]. The training compound will belong to the applicability domain of the model if the leverage (h_i) is predicted less than the critical value (h^*) [27].

2.3. Molecular docking

In drug discovery, molecular docking is a key tool for predicting ligand binding ability to target proteins [28]. Moreover, molecular docking was revealed early to be a valuable tool to understand and predict molecular recognition in drug discovery [29].

In the current study, before molecular docking calculations, all thirty-one compounds were fully optimized using standard DFT calculations at the B3LYP/TZP level (vide supra). Therefore, the molecular docking technique was utilized to determine the best poses of the optimized compounds within the four different cancer cell lines; ovarian cancer A2780 (PDB ID: 4LXZ Resolution: 1.85 Å) [30], breast cancer MDA-MB 231 (PDB ID: 4GL7 Resolution: 3.90 Å) [31], human lung adenocarcinoma A549 (PDB ID: 1M17 Resolution: 2.60 Å) [32] and HCT116 colorectal cancer cell lines with the human cyclin-dependent kinase 2 (CDK2) (LPDB ID: 6GUE Resolution: 1.99 Å) [33].

The PDB files were obtained from the Protein Data Bank (RCSB) [34] as shown on Figure 2. Virtual Molegro Docker MVD 2019.7.0 release software [35] was used to perform the docking simulations.

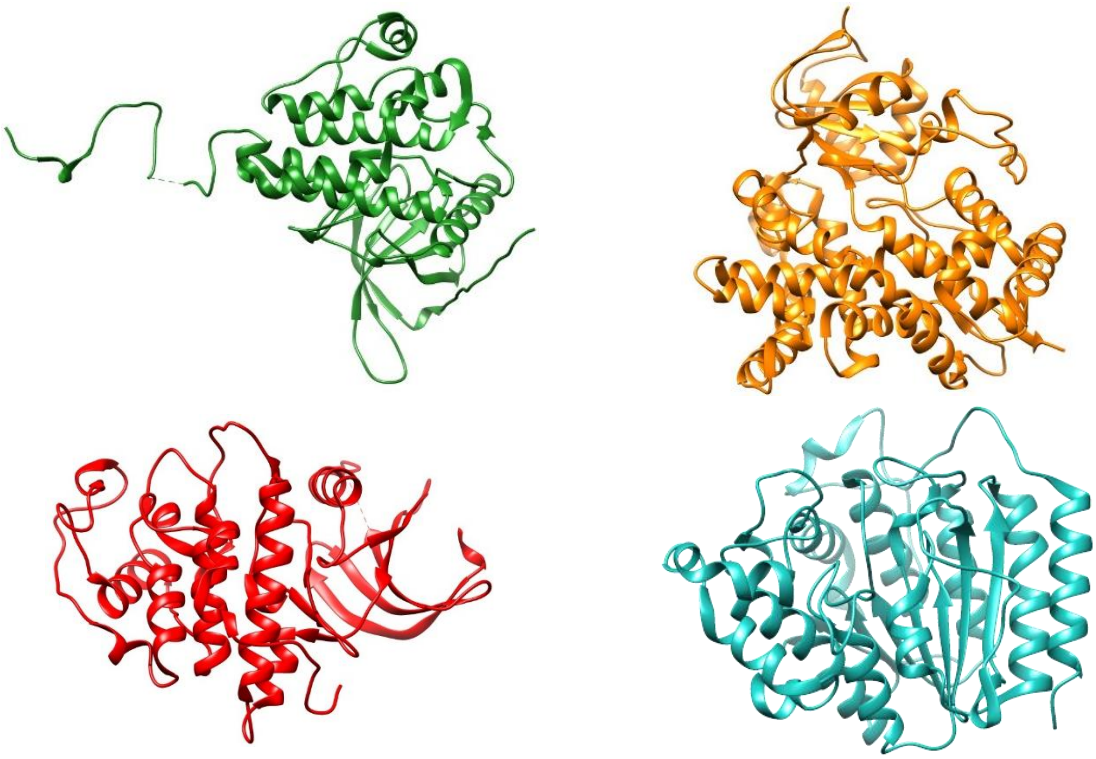


Figure 2: 3D structures of proteins: tyrosine kinase (PDB ID: 1M17) (green), placental aromatase (PDB ID: 4GL7) (orange), cyclin-dependent kinase 2 (PDB ID: 6GUE) (red), and histone deacetylase 2 (PDB ID: 4LXZ) (bleu).

Initially, the four protein structures (PDB codes: 4LXZ, 4GL7, 6GUE, and 1M17) have been preprocessed and refined, involving removing water molecules, ions, and other heteroatoms. The Autogrid algorithm as implemented in Molegro, was used to identify the active site that will be used to evaluate the interaction energy between the ligands and the proteins as reported in Table 2.

Table 2: Coordinates of the x, y, and z centers of the active site within the proteins dimensioned with volume and surface

Target	x- center (Dimension)	y-center (Dimension)	z-center (Dimension)	Surface A^2	Volume A^3
4LXZ	25.790	-16.070	1.001	85.760	23.040
4GL7	86.285	51.466	43.909	706.501	327.401
6GUE	-4.807	-21.898	21.521	4.161	160.701
1M17	27.701	3.661	49.998	842.241	261.120

The dimensional parameters of Table 2 are chosen to allow large size of active sites within considered proteins, and to maximize the reliability and efficiency of the docking simulations.

2.4. Molecular Dynamics (MD)

The best conformations obtained from the molecular docking are used as starting points for molecular dynamics (MD) simulations. CHARMM-36-2019 release force field [36] implemented in the Gromacs-GPU package [37] was used to assess the stability of the best-docked protein ligands.

SwissParam serves as a user-free platform designed for the generation of crucial parameter files (ITP files) required in molecular dynamics simulations, particularly when employing the CHARMM force field for small organic molecules. To utilize this tool, it is essential to submit the molecule in SYBYL molecule (mol2) format on the SwissParam website.

The MD technique was conducted at 1 atm and 310.15 K in a simulation box. A cubic box, maintaining a buffer distance of 11.5 angstroms, accommodated the models of protein and ligand complexes. The protein complexes underwent solvation utilizing the TIP3P (transferable intermolecular potential 3P) water model [38]. To neutralize the system, chloride anions (Cl^-) and sodium cations (Na^+) were introduced. The energy of the system was minimized using the steepest descent method for 50.000 steps, followed by a 2 nano-second equilibration in both the canonical (NVT) and isothermal isobaric (NPT) ensembles using the Berendson thermostat [39], and Parrinello-Rahman pressure coupling at 1.0 bar [40] with a time step of 2 fs. It is noteworthy that a comprehensive approach was thoughtfully carried out to ensure the reliability and accuracy of our simulations. Indeed, prior to starting the simulations, an in-depth analysis of all assumed NVT and NPT parameters was performed. This assessment enabled us to establish a precise configuration, forming the essential foundation for the reliable execution of the simulation process.

While minor pressure imbalances may have limited effects in the majority of contemporary MD simulations, we anticipate that their significance will increase as we delve into investigations involving progressively larger systems, especially when coupled with anisotropic pressure control. Our system likely aligns with the typical scenarios encountered in current MD simulations, which implies that the barostat parameters in GROMACS are accurate.

The MD simulations were then performed for 200 ns, and the stability of protein-ligand complexes was assessed using the root mean square deviation (RMSD), the root mean square fluctuation (RMSF), and the radius of gyration (Rg).

Extensive analysis of dynamics simulations was carried out using the principal component analysis (PCA), the solvent accessible surface area (SASA) [41, 42] and Free Energy Surface (FES). PCA is a statistical procedure to assess the collective motion in biological macromolecules during the course of molecular dynamic simulations. The latter, are performed with reduced dimensions of the dataset keeping the critical information which is characterized by the eigenvectors. Moreover, the SASA approach serves as a critical parameter for assessing the structural compactness of proteins and their ligand complexes during MD simulations.

The PCA, SASA, and FES of the systems (ligand-receptor) were generated with gmx_sasa, gmx_covar, and gmx_anaeig command line of GROMACS software respectively, and used as an index of the stability of the complexes. The probability distribution of the SASA over 200 ns of the MD was obtained with gmx_analyze. The complexes were then determined as the eigenvectors of the mass-weighted covariance matrix of the backbone of the protein.

2.5. ADMET and drug-likeness prediction

The oral bioavailability and toxicity are also assessed through computational ADMET (absorption, distribution, metabolism, excretion, and toxicity) analysis [43, 44, 45, 46]. Indeed, the pharmacokinetic factors analysis was performed to examine the drug-likeness of the best-fitted ligands obtained by molecular docking. Furthermore, the pharmacokinetics and toxicity profile of the new twelve designed compounds were characterized in silico using the ADMET Predictor™ cloud version [47] and SwissADME [48]. ADMET Predictor™ is a statistical-based QSAR software allowing in silico pharmacokinetics and toxicological assessment with experimental standard ranges (https://www.simulations-plus.com/software/admetpredictor/).

Indeed, lipophilicity or octanol-water partition coefficient LogP, 2D polar surface area (PSA). Gastrointestinal absorption (GI), molar solubility in water LogS (mg/mL), blood-brain barrier (BBB) permeability, human gastrointestinal absorption (GI), and other factors that are expressed in log units, and the toxicity of the potential new inhibitors are predicted. Furthermore, bioavailability score are among the pharmacokinetic characteristics considered to assess drug-likeness, according to the Lipinski's rule of five.

Moreover, according to ADMET Predictor™ analysis, absorption risk (Absn_Risk) with a score in the range of 0-8 indicating the number of potential oral absorption problems a compound is likely to have. Also, the full ADMET risk parameter (ADMET_Risk) with a

score range of 0-22 indicates the number of potential ADMET problems a compound might have. The P-glycoprotein substrate and inhibitor ability are predicted by Pgp_Substr and Pgp_Inh parameters, respectively.

Metabolism is predicted based on the CYP models for substrate or inhibition. These latter were calculated and checked for compliance with their standard ranges.

3. Results and discussion

3.1 QSAR analysis

The 2D-QSAR technique was utilized on the considered dataset of thirty-one hybrids of phenylsulfonyl furoxan and phenstatin derivatives, which revealed recently significant multi-target anticancer inhibition activity [12]. Our in silico study involves different cell lines; A2780, MDA-MB-231, HCT116, and A549 related to the four tumor cancer types i.e. breast, colorectal, lung, and ovarian diseases, respectively.

Therefore, 2D-QSAR model was developed using multiple linear regression (MLR) technique, and the statistical parameters were computed using the R-studio software [21]. The developed equations for each cell line with six and seven molecular descriptors showed significant correlations between the predicted pIC50 values of biological activity, and the molecular descriptor, as given by equations (4–7) for each cell line, are obtained as follows:

Cell line (A2780):

$$pIC50 = 0.004 *MW – 0.77 *nX + 12.3 *MSD – 3.85 *MATS1m -0.0173 *DISPm +9.28 *E2v – 0.379 \quad (4)$$

Cell line (MDA-MB 231):

$$pIC50 = 0.0025*MW – 0.807*nX + 9.48*MSD – 3.41*MATS1m +1.91*MATS6m +7.78*E2v + 0.09 \quad (5)$$

Cell line (HCT116):

$$pIC50 = 0.0023*MW - 0.435*nX - 0.63*GATS4m - 0.0358*DISPm -1.53*Mor13u - 1.6*E3m + 9.16*E2v +2.41 \quad (6)$$

Cell line (A549):

$$pIC50 = 0.0025*MW – 0.19*nX +2.55*MSD – 2.39*MATS1m + 2.1*MATS6m + 7.44 *E2v + 5.26 \quad (7)$$

where the molecular descriptors (MW, nX, MSD, MATS1m, MATS6m, E2v) are defined as follows:

Molecular descriptor	**Description**
MW	: (Molecular Weight) the sum of atomic weights in a molecule.
nX	: (Number of Atoms of Type X) count of atoms of a specific type X.

MSD : (Mean Squared Deviation) a statistical measure of dispersion in a set of values.

MATS1m : (Moran Autocorrelation of lag 1 weighted by atomic masses) is au autocorrelation of atomic masses with a lag of 1.

MATS6m : (Moran Autocorrelation of lag 6 weighted by atomic mass) is an autocorrelation of atomic masses with a lag of 6.

E2v : (Balaban-like index from electronegativity weighted by van der Waals volumes) combines electronegativity and van der Waals volumes for electronic and spatial characteristics.

The results of statistical parameters (R^2, R^2_{adj}, MSE, Q^2, R^2_{test}) for the four cell lines are reported in Table 3.

Table 3: Statistical parameters of the built 2D-QASR models of the four cell lines

	A2780	MDA-MB-231	HCT-116	A549
R^2	0.85	0.74	0.82	0.75
R^2_{adj}	0.79	0.64	0.74	0.661
MSE	0.094	0.097	0.110	0.011
Validation LOOCV				
Q^2	0.690	0.650	0.610	0.608
External Validation				
R^2_{test}	0.761	0.730	0.710	0.701
Y-randomization				
R^2_{rand}	0.160	0.081	0.110	0.190

It is noteworthy that the predicted biological activities (pIC50) of the thirty-one compounds (Table 4), are in good agreement with those in vitro observed [12], which in turn indicate that the MLR models are satisfactorily reliable.

Table 4: Experimental and predicted activities of the thirty-one compounds dataset of all cell lines. (*) are test set.

pIC50	Experimental				Predicted			
Mol_ID	A2780	MDA-MB-231	HCT-116	A549	A2780	MDA-MB-231	HCT-116	A549
1a*	6.92	6.29	5.77	5.77	7.11	6.36	5.7	5.88
1b	5.35	4.91	4.25	4.49	5.45	4.37	4.26	4.18

1c	5.53	4.33	4.45	4.28	5.54	4.47	4.34	4.23
1e	5.43	4.38	4.58	4.58	5.44	4.67	4.35	4.45
1f	5.49	4.49	4.32	4.32	5.59	4.44	4.18	4.25
1g	6.34	5.25	4.9	4.9	6.26	5.15	4.83	4.89
1h*	5.94	4.68	4.44	4.44	5.82	4.67	4.22	4.43
1i	6.77	5.92	4.91	4.91	6.46	5.74	4.97	5.00
1j	4.91	3.88	3.84	3.84	5.19	4.5	4.42	4.54
2a	7.46	6.08	5.99	5.99	7.29	6.2	5.84	5.69
2b*	6.28	5.14	4.44	4.44	6.8	5.57	4.26	5.12
2c	6.54	5.77	5.62	5.62	6.74	5.54	5.63	5.6
2d	6.82	5.68	5.8	5.8	6.8	5.64	5.96	5.7
2e	7.39	5.96	6.08	6.08	7.34	5.96	6.06	5.92
2f	6.82	5.27	5.54	5.54	7.26	5.72	5.55	5.64
2g	7.39	5.82	5.37	5.37	7.24	5.75	5.43	5.66
2h*	7.92	6.82	6.96	6.96	7.38	6.25	6.85	6.06
2i*	6.92	5.44	6.02	6.02	6.85	5.38	6.21	5.91
2j	7.54	6.25	5.96	5.96	7.04	5.67	5.65	5.62
3a	6.77	6.14	5.49	5.49	6.78	5.76	5.14	5.14
3b	6.49	5.32	5.29	5.29	6.04	4.95	4.76	4.9
3c*	6.28	5.14	4.44	4.44	6.21	4.89	4.58	4.05
3d	6.46	5.21	5.04	5.04	6.1	5.06	4.88	5.00
3e	6.62	4.92	4.77	4.77	6.53	5.26	4.85	5.11
3f	6.26	4.95	4.66	4.66	6.75	5.16	5.22	4.97
3g	6.47	5.23	4.92	4.92	6.67	5.15	4.99	4.95
3h	6.82	5.18	5.33	5.33	6.69	5.34	5.33	5.15
3i	6.36	5.42	4.64	4.64	6.89	5.93	4.86	5.33
3j*	6.59	5.57	5.21	5.21	6.8	5.96	5.36	5.83
4*	6.06	4.73	4.57	4.57	5.94	4.73	4.62	4.57
5	6.04	4.94	4.64	4.64	5.97	4.86	4.86	4.51

This high correlation between predicted pIC50 and observed biological activity is also illustrated on Figure 3. Indeed, the scatter William's plots of observed biological activities vs. predicted MLR.pIC50 data values of pIC50 for the 2D-QSAR models of A2780, MDA-MB-231, HCT-116, and A549 cell lines, sustain the high R^2 correlation approach and show that the selected descriptors are well linearly correlated with the observed anticancer biological (pIC50) activity values.

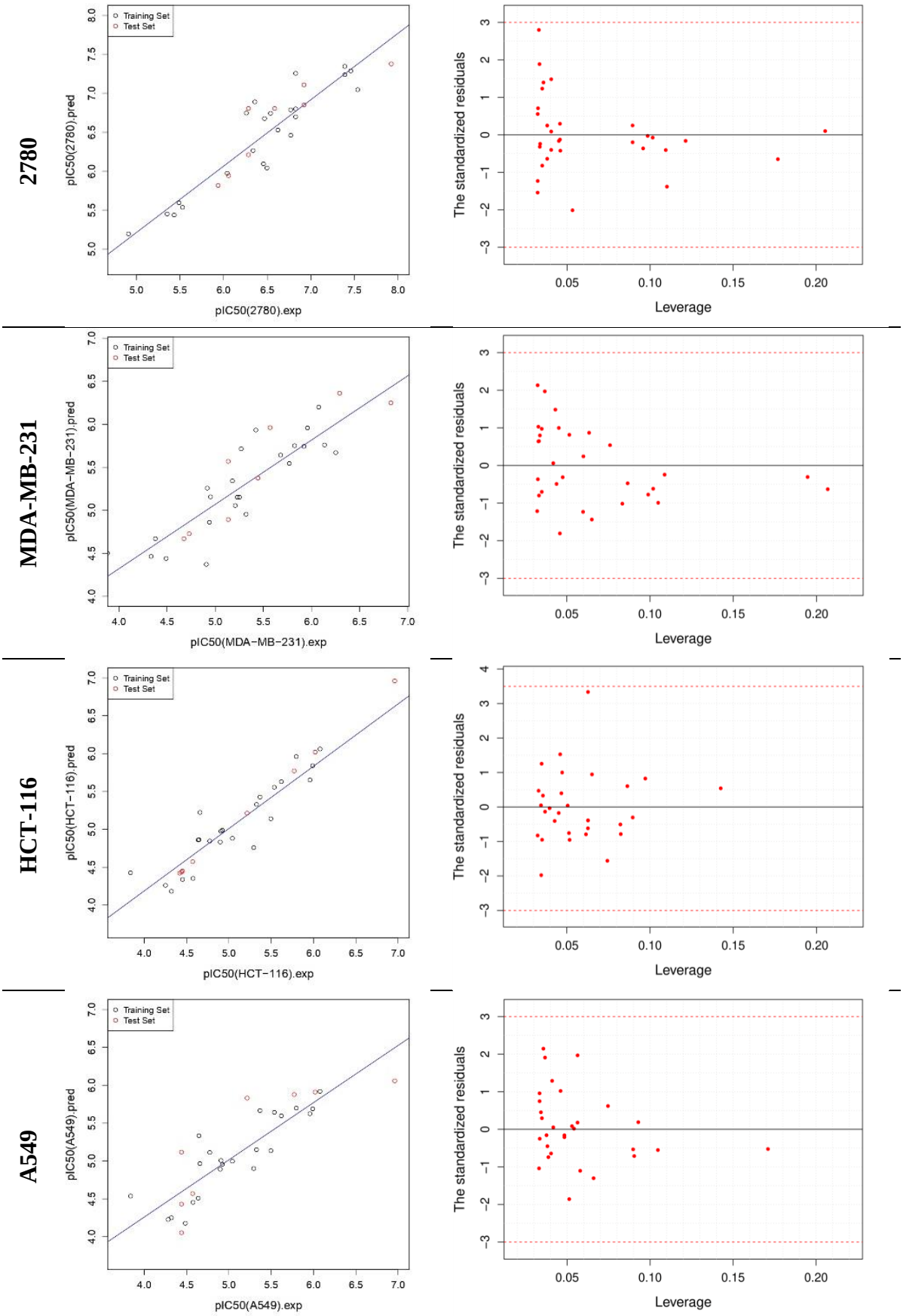


Figure 3: Scatter plots of observed vs. predicted values of pIC50 for 2D-QSAR models and William's plots of A2780, MDA-MB-231, HCT-116, and A549 cell lines

Furthermore, the coefficient of determination R^2 values (Table 3), obtained from the multiple linear regression (MLR) analysis, ranged from 0.74 to 0.85 for A2780, MDA-MB-231, HCT116, and A549 cell lines (Table 3), indicating that the best-fitting MLR constructed model accounts from 74 to 85 percent of the experimental pIC50 values.

Moreover, the adjusted squared (R^2_{adj}) values provide an accurate estimate of the MLR model's fit by taking into account the number of predicted variables in the model. The values obtained for A2780, MDA-MB-231, HCT116, and A549 cell lines are 0.79, 0.64, 0.69, and 0.66, respectively, which indicate a good fit for the models.

The MSE value measures the average squared difference between the predicted and actual values of the response variable. The lower the MSE value, the better fit of the model to the data. In our case, the MSE values obtained for A2780, MDA-MB-231, HCT116, and A549 are 0.094, 0.097, 0.110, and 0.011, respectively. These results indicate that the MLR models have a relatively low error rate and can predict satisfactorily the pIC50 values of new compounds. For internal validation, the squared leave-one-out cross-validated (LOOCV) Q^2 values obtained for the MLR models for these cell lines (Table 3), are all greater than 0.6, sustaining the reliability of the 2D-QSAR models.

Also, external validation is an important step in the 2D-QSAR model development as it provides an assessment of how well the model will perform on new and unseen data. The R^2_{test} values obtained for the MLR models for A2780, MDA-MB-231, HCT116, and A549 cell lines (Table 3), are higher than 0.70, which indicates a good fit of the validation data. Besides, this means that the 2D-QSAR model is satisfactory and reliable for predicting the activity of new multitargeting compounds relative to such cell lines.

It is of interest that such external validation (R^2_{test}), including training and testing series, should be conducted randomly to confirm the robustness of the models [15]. Indeed, as shown in Table 3, the Y-randomization (R^2_{rand}) approach shows low-test values, ranging from 0.08 to 0.19, indicating that the original R^2_{test} values obtained from the training model on the unpermuted data are significantly higher than the obtained R^2_{test} values from the training model on the permuted data.

One can conclude that the 2D-QSAR model is built on a real relationship between the molecular descriptors and the activity values (Eq. 4-7), instead of being founded on random connections.

Turning back to the Williams plot Figure 3), which displays standardized residual values against leverage values, is an important tool used to determine the applicability domain (AD) of a building 2D-QSAR model. This graphical analysis helps to identify and exclude outlier compounds that could compromise the model's reliability and accuracy. Indeed, Figure 3 shows that the 2D-QSAR model was constructed using 6 and 7 descriptors for a training set of 23 compounds, with a normalized residue limit of 3.

The leverage h_i (Eq. 2) values were determined, and it was found that all the compounds in the training and test sets were within the domain of applicability since their leverage values were lower than the critical limit (h^*= 3(6 + 1)/32 = 0.67), and their standardized residuals were within the range of ±3.

3.2. Molecular Docking

To validate the molecular docking analysis, it is crucial to carry out a re-docking test, keeping the protein structure fixed, and the ligand re-docked into its crystal-binding pocket. The comparison is based on the root mean square deviation (RMSD) between the docked and crystal structure poses. The docking technique is considered satisfactory if the RMSD range does not exceed 2Å [49, 50]. The re-docking experiment was performed using Molegro software for each of the four protein structures as shown on Figure 4. Indeed, the best pose obtained gave RMSD values of 0.69, 0.90, 1.80, and 1.57 Å for 4GL7, 6GUE, 1M17, and 4LXZ targets, respectively. These results suggest that the re-docking test reproduces satisfactorily the active site of the crystal reference ligand poses within the four proteins.

.

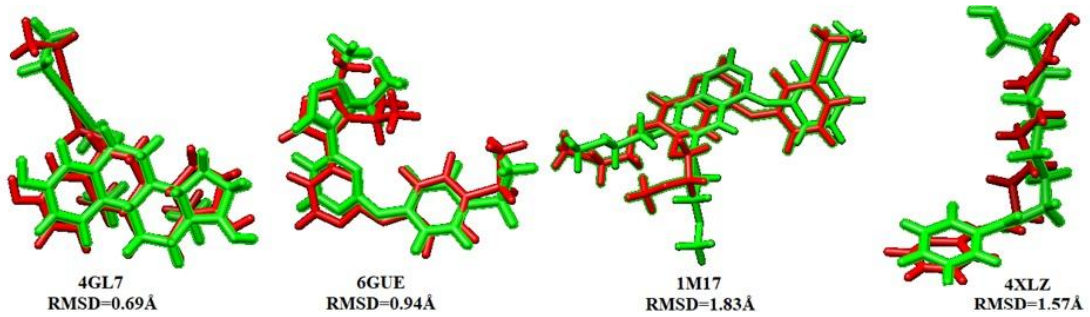


Figure 4: Conformational relationship between the pose and its reference ligand in the inhibitor pocket (Red = Original, Green = Docked)

In the second step, molecular docking analysis was performed to investigate the best residues interactions that contribute to the anti-cancer activity of the multitargeting compounds under consideration. The results for the twelve ligand-receptor complexes involving the best-

predicted inhibitors (**2e**, **2i**, **2j**, **2h**, **3c**, **3h**, **3i**) associated with the four cell lines (4GL7, 6GUE, 1M17, 4XL7), including the highest docking score (kcal/mol), hydrogen bond energy (kcal/mol), and amino acid steric interactions, are reported in Table 5.

Table 5: Molecular docking results of docking scores and hydrogen bonds (kcal/mol) for the three best ligand-receptor compounds and the native ligands.

Receptor	Ligand	Mol Docking Score (kcal/mol)	Hydrogen Bond (kcal/mol)
4GL7	Native 0XJ_601	-146.225	-5.649
	2j	-174.681	-4.939
	3i	-169.688	-3.294
	2e	-163.814	-0.684
6GUE	Native FB8_301	-132.743	-3.42109
	3c	-146.157	0
	2e	-142.190	0
	2j	-141.930	-0.417
1M17	Native-AQ4	-111.383	-2.718
	3h	-126.561	-1.172
	2i	-125.908	-2.475
	2h	-123.204	-1.812
4LXZ	Native SHH_407	-102.362	-1.576
	2j	-140.314	-0.483
	2i	-126.736	-3.827
	2e	-125.312	0

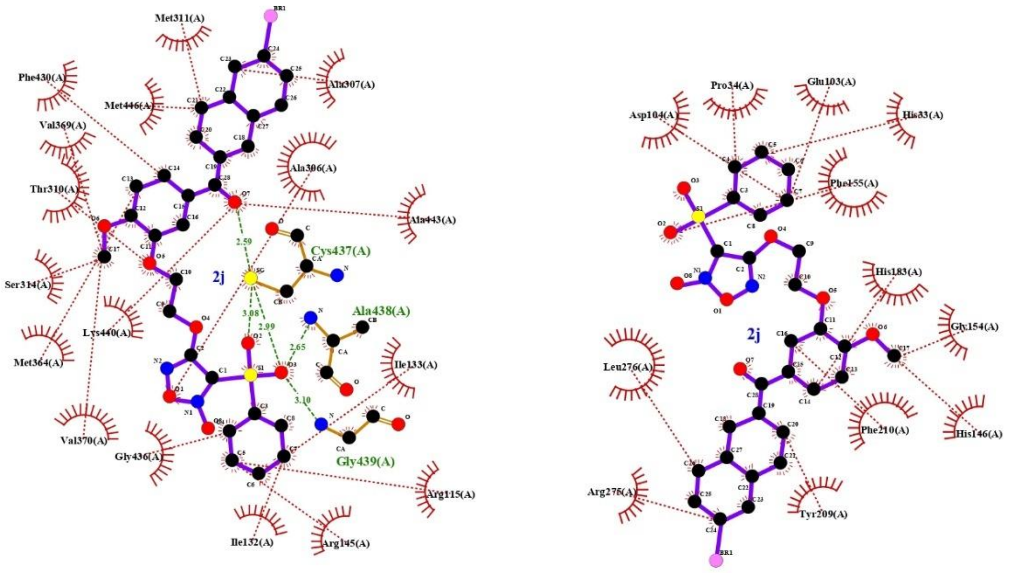
On the Figure 5, are shown the potent 2-D interactions between the best ligands scoring and active site residues of the four target complexes i.e., 2j-4GL7, 3c-6GUE, 2h-1M17, 2j-4LXZ. It's noteworthy that multitargeting simulations aimed to predict how well the compounds might best bind to the protein. Indeed, according to the LigPlot analysis [51] and the results of Table 5.

These tables compare the differences between the native ligands and other ligands for each receptor in terms of molecular docking scores and hydrogen bond energies. Negative values indicate that the native ligand has a higher docking score or higher hydrogen bond energy

than the compared ligands **2j**, **3i**, **3c**, **2e**, **3h**, **2i**, and **2h**, suggesting a potentially stronger binding affinity or stability with these ligands.

The predicted interactions between ligands and active sites, show that the 2j-4GL7 complex, exhibits the best scoring energy (-174.681 kcal/mol) and is stabilized mainly by three strong hydrogen bond interactions with Cys437, Ala438, and Gly439 residues, totalizing H-bonds of -4.94 kcal/mol (3.1 Å). Moreover, the second-best scoring 3c-6GUE complex (-146.16 kcal/mol), is stabilized mainly by steric interactions with no H-bond interactions, which further contribute strongly to the cohesive environment. For the third 1M17 target, the best-observed in-vivo evaluated **3h** compound [12], show slightly higher docking score kcal/mol) than the interacting **2i** and **2h** ligands scores (-126.5 vs. -125.9 and -123.2 kcal/mom).

Similarly, for the best observed antitumoral in-vivo **3h** scores, the molecular docking simulations with the 1M17 protein Figure 5), show two strong H-bond interactions with GLU722 and PHE699 amino acids, totalizing -1.812 kcal/mol. Steric interactions LEU763-723, ASP831, GLY833, GLU738, GLY700 and ILE735 residues, were also seen and are almost identical to those found in the complex produced by compound **2j**. In contrast, the complex (2j-4LXZ) exhibits predominantly steric interactions with relatively weak hydrogen bonds within the complex.



4GL7 4LXZ

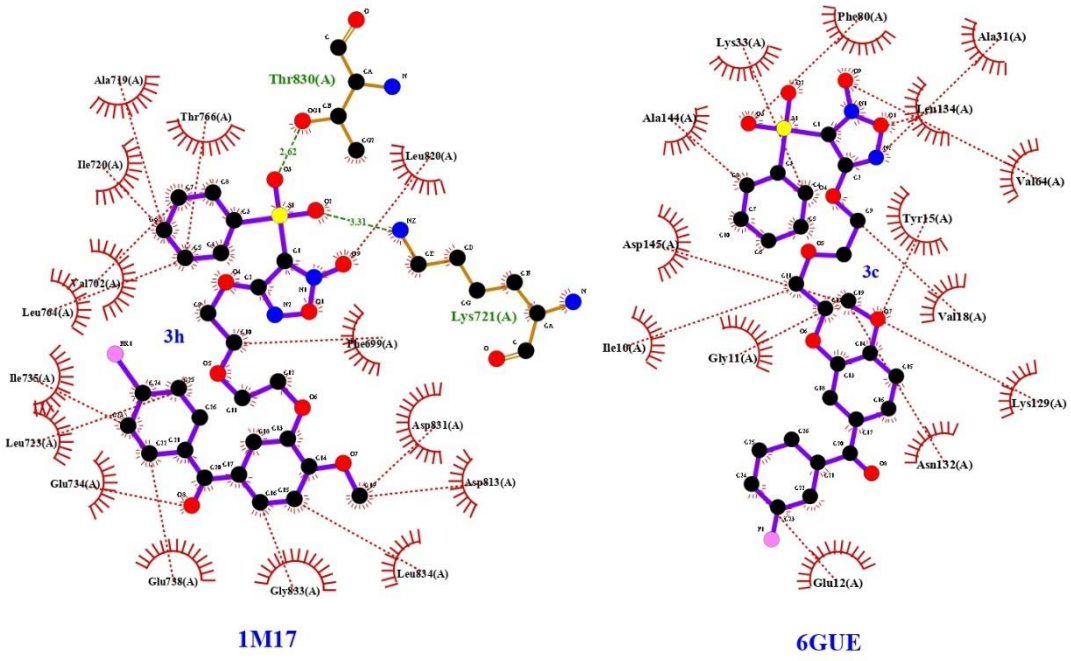


Figure 5**:** 2-D interactions between the active site residues of the targets and best docking score ligands (2j-4GL7, 3c-6GUE, 3h-1M17, and 2j-4LXZ)

In conclusion, both hydrogen bond and steric interactions play a crucial role in stabilizing the compounds and enhancing their biological activity. The inhibitory impact of **2j**, **3c**, **2h** and **3h** against multi-target cancers (ovarian, breast, colorectal, and lung) was strengthened by interactions with specific active sites of the targeted protein.

3.3. Molecular dynamics (MD)

Nowadays, the MD simulation is commonly used in various in silico biological applications to simulate the structural and physiological perturbations with real-time mobility of the complexes. The analysis of trajectory was performed using GROMACS utilities, and are obtained at the end of the simulation providing useful and detailed information on the stability of the proteins alones, and ligand-receptor complexes, as well as their molecular interactions.

The preliminary assessment of MD simulation data was carried out using the root-mean-square deviations (RMSD) of the backbone of all systems with respect to their initial coordinates to explore the stability of the four proteins alones, and their complexes. Furthermore, the root mean square fluctuation (RMSF), radius of gyration (Rg), the principal component analysis (PCA), and solvent accessible surface area (SASA) as well as free energy

surface (FES) approaches of the systems backbone, were carried out to explore the stability of high scoring multitargeting ligands within the four 4XLZ, 1M17, 4GL7, and GUE targets.

To analyze the stability of the proteins and their complexes, the resultant trajectories of the RMSD, RMSF, and Rg were further analyzed using g_rmsd, g_rmsf, and g_gyrate GROMACS utilities, respectively, and are depicted on figures 6, 7, and 8, respectively. On Figure 6 are shown the RMSD of native proteins (4XLZ, 1M17, 4GL7, and GUE) and their twelve complexes (2e-4XLZ, 2i-4XLZ, 2j-4XLZ, 2h-1M17, 2i-1M17, 3c-1M17, 2e-4GL7, 2j-4GL7, 3i-4GL7, 3c-6GUE, 2j-6GUE, 2e-6GUE), exhibiting the best docking score varying from 0.15 to 1.6 nm, and reaching stability after the first 30 ns simulation.

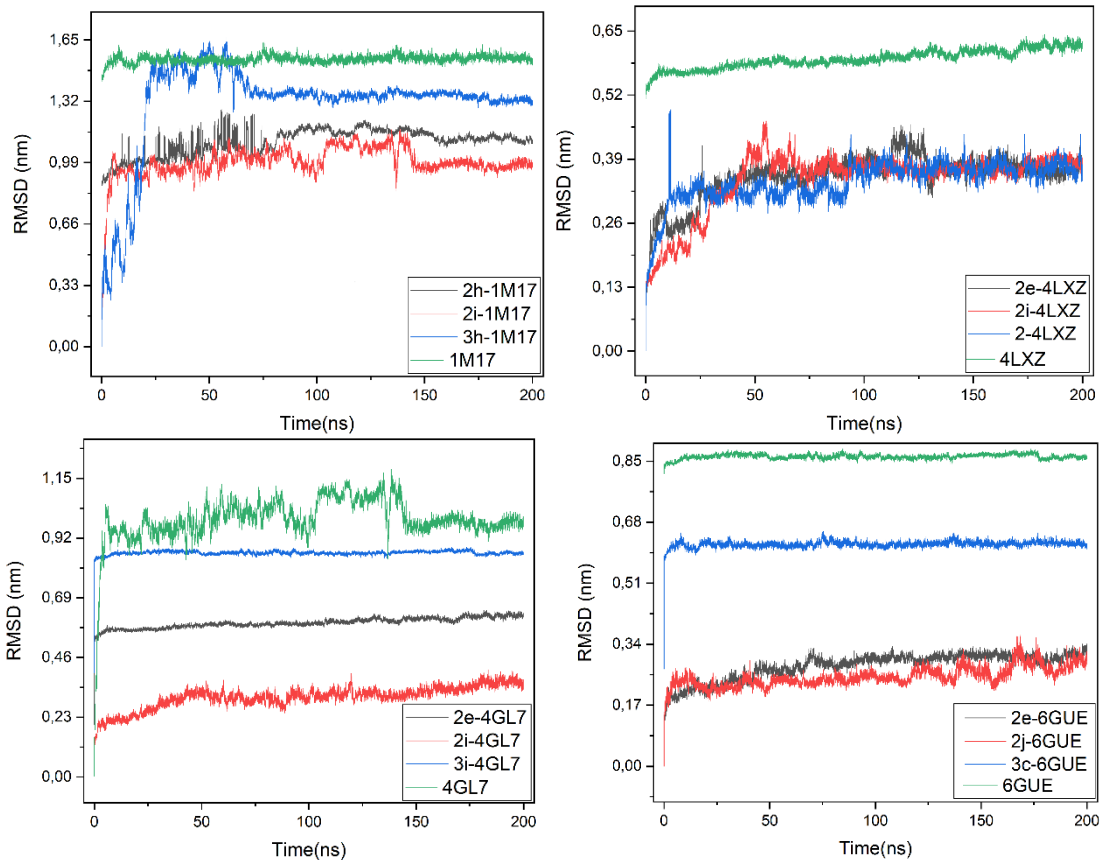


Figure 6: RMSD of the four proteins alone and their complexes over the 200 ns of simulation.

It is noteworthy that the RMSD values for the free proteins are significantly higher than those of their complexes (protein+ligand) as shown on Figure 6, suggesting that the latter are more stable. Moreover, the RMSD for the protein 1M17 is significantly high (16 Å), however, such a threshold value is not unexpected, regarding previous related *in silico* studies on the native protein 1M17 [57,58,60]. Indeed, dynamics simulation at the same force fields but with a

time scale of 0-100 ns, have led to close or beyond 16 Å values [60]. Moreover, the inhibition activity is related to many factors including the stability of complexes.

Notably, the RMSD analysis of the complexes 2i-1M17 and 2h-1M17 relative to the 1M17 protein, shows that such systems are best stabilized around 1.05 nm, in opposite to the 3h-1M17 congener, which shows some variability in the initial time up to 200 ns, and then stabilize with average values of 1.40 nm.

For the 4XLZ protein, the three high scores interacting **2e**, **2i**, and **2j** ligands, the RMSD values are mainly below 0.4 nm, which indicate satisfactory stability of their complexes during the simulation. Concerning the two 4GL7 and 6GUE targets, their interactions with scoring ligands i.e., **2e/2j/3i** and **2e/2j/3c**, respectively, show that for the best docking score 2j-4GL7 complex, it quickly gets stabilized around 0.34 nm, while for the second target (6GUE), the two **2j** and **2e** complexes, get systematically stabilized below 0.3 nm, in opposite to the less stable **3c** congener (~0.65 nm) [52].

One can note that high fluctuation of the RMSD values e.g., for the 1M17 complexes, is not unexpected, regarding MD simulation and docking techniques which are often considered satisfactory if the RMSD deviation range must not exceed 2Å [50]. However, it is noteworthy that such cutoffs of 2Å can be used to select representative structures in molecular dynamics simulations. Indeed, the interpretation of the RMSD/RMSF/Rg mean values is context-dependent and should be used alongside other analyses and experimental data to understand a molecular system [53-60].

For example, in simulations of well-folded globular proteins, an RMSD of 1-2 Å (0.1-0.2 nm) is often considered of a good indicator. For simulations of intrinsically disordered or partially unfolded proteins, higher RMSDs can be observed, typically of 2-5 Å (0.2-0.5 nm) or more [55, 58, 59]. These dynamics simulations data show that all complexes, exhibiting best docking score, have significant affinity for residues present in the active site, suggesting their high stability under physiological conditions.

The average deviation of the atom in the simulation from a reference position was shown by the RMSF analysis as depicted on Figure 7. Moreover, the RMSF values represent the thermodynamic stability and rate of mobility of all residues. Notably, the RMSF of the complexes is roughly the same, with low values (less than 2.5 nm) indicating that the ligands

do not undergo significant conformational changes over time, despite the occurrence of different amino acids in the four systems as stated in previous work [60]. However, the flexible residues in the ligand-binding areas slightly changed upon recognition of ligands to accommodate them and maintain equilibrium.

For the four 1M17, 4LXZ, 4GL7, and 6GUE proteins, the predicted average RMSF values for their 2h/2i/3h, 2j/2e/2i, 2e/2j/3i, and 2e/2j/3c complexes, are 0.25, 0.18, 0.15, and 0.12 nm, respectively, and show that active site residues were not considerably perturbed upon binding of the ligands [57, 61].

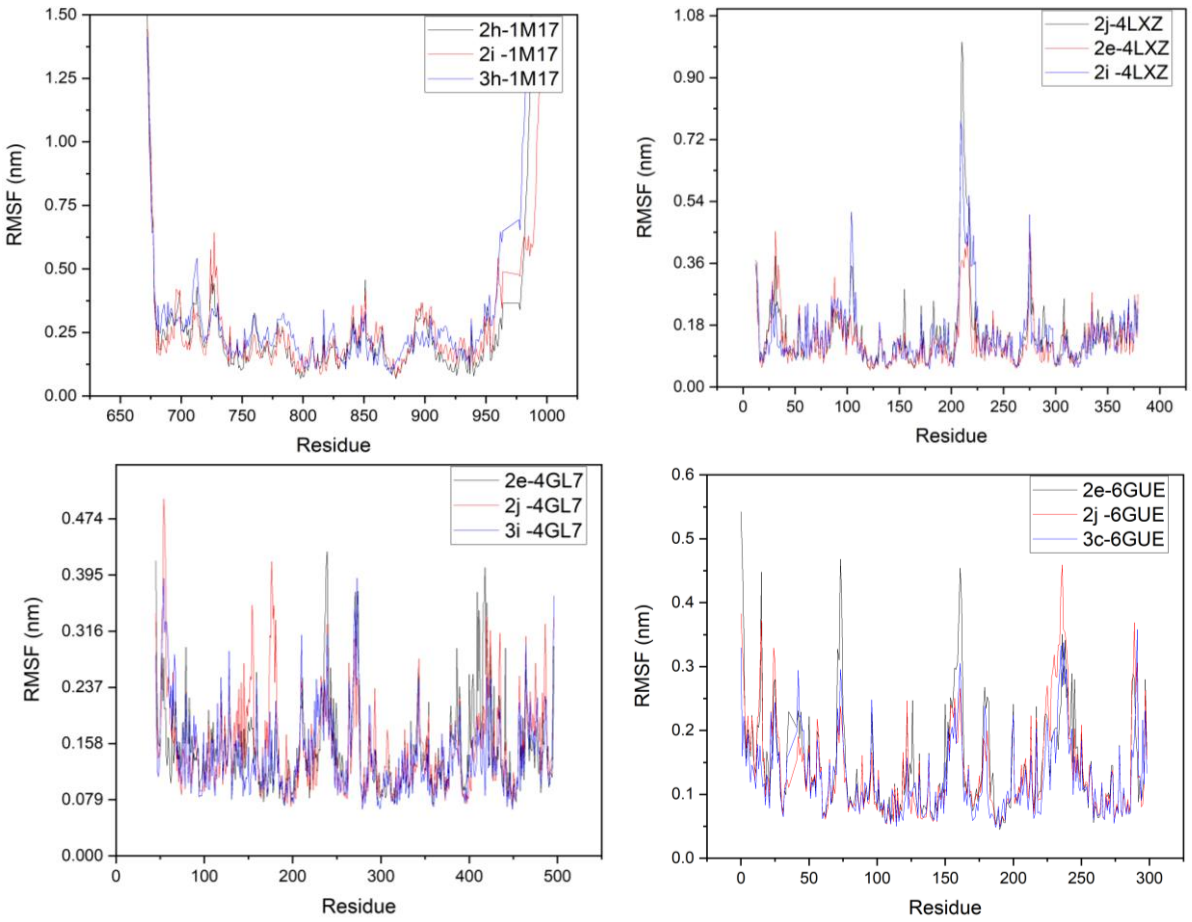


Figure 7: RMSF of the proteins 1M17, 1LXZ, 4GL7, and 6GUE with ligands

The RMSD and RMSF scores obtained indicate the greater stability of the best docking score systems compared to the four non-complexed (free) proteins.

Moreover, according to the MD simulation data, the twelve protein-ligand complexes remained stable over 200 ns. This is supported by the R_g calculation (Figure 8), which

remained consistent with the average value being 2.25, 2.04, 2.18, and 2.10 nm, for the four different systems i.e., 4GL7, 6GUE, 1M17, and 4LXZ, respectively. In particular, the R_g analysis indicates both molecule's stability, structure dimensions, and compactness. In general, a stably folded protein tends to maintain a relatively less variation in R_g value, which determines its dynamic stability. This small variation occurring in R_g value between 2.1 and 2.5 nm Figure 8), shows that the compactness of ligand-receptor complexes is relatively stable [58].

Furthermore, these results of R_g values indicate that the high-scoring ligands remained stably bound to the proteins without inducing any significant alterations in their structure.

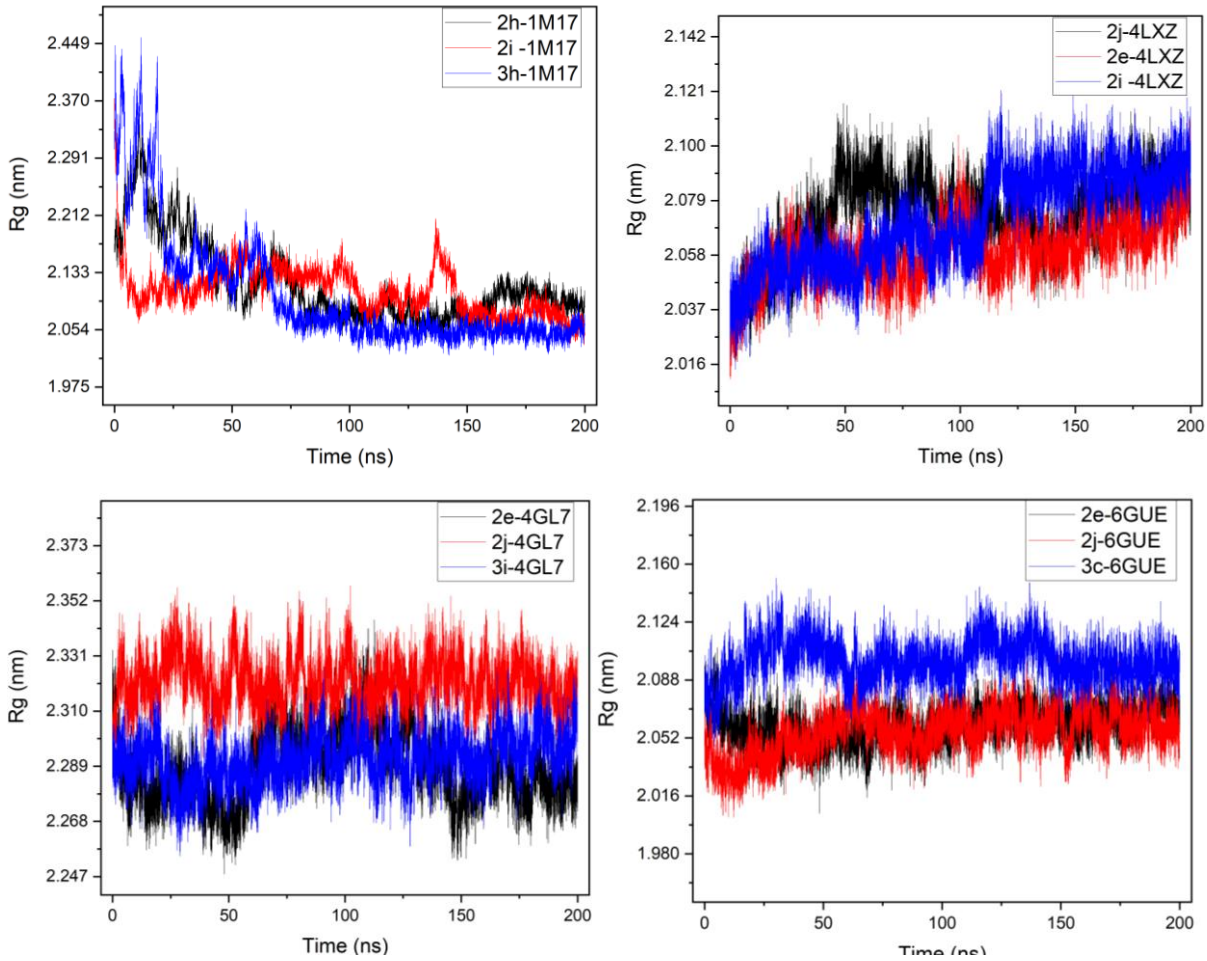


Figure 8: R_g of proteins and ligands

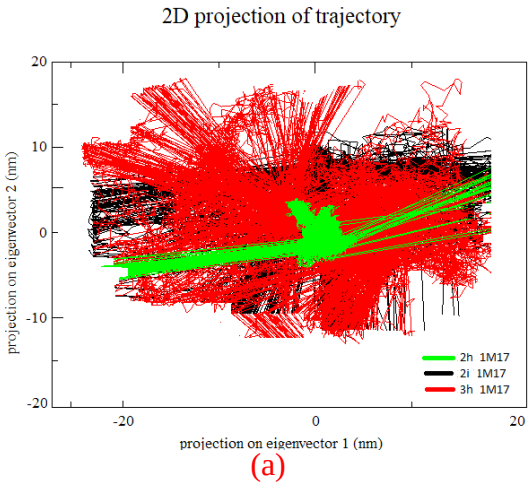
It is noteworthy that in our investigation of complexes' stability, the evaluation of RMSE parameters during dynamics simulation plays a crucial role in providing insights into the system's behavior. Indeed, the obtained RMSE parameters using the simulations were in line

with the literature [52, 56, 59, 60, 63], and are consistent with the high stability of the complexes as previously stated [27]

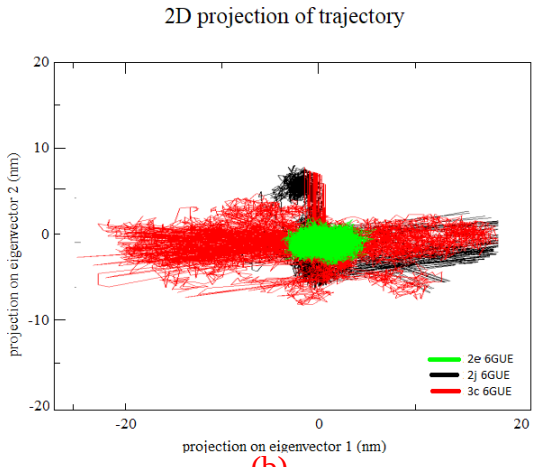
Principal Component Analysis PCA

To assess in-depth the flexibility of the ligands within the four targets, scatter plot of principal component analysis (PCA), was carried out by 2D projecting the eigenvectors of the twelve complexes associated with the proteins: 1M17, 4LXZ, 4GL7, and 6GUE as depicted on figure 9. From the 2D projection, a complex with a stable cluster occupying less phase space represents a stable complex, and in the opposite, a non-stable cluster is found to occupy more scattered space conformation, which represents a less stable complex [55].

Therefore, both Tyrosine kinase 2h-1M17 figure 9a), Cyclin-dependent kinase 2e-6GUE and 2j-6GUE figure 9b), the Placental aromatase 2e-4GL7 and 3i-4GL7 figure 9c), as well as the Histone deacetylase 2i-4LXZ and 2e-4LXZ complexes (figure 9d), are predicted to occupy significantly compactness and less conformational space compared to their analogous, indicating more structural stability of their complexes. These findings indicated that selected 2h, 2e, 2i, and 3i, ligands are proficient inhibitors of targets due to their stable complex formation characteristics.



(a)



(b)

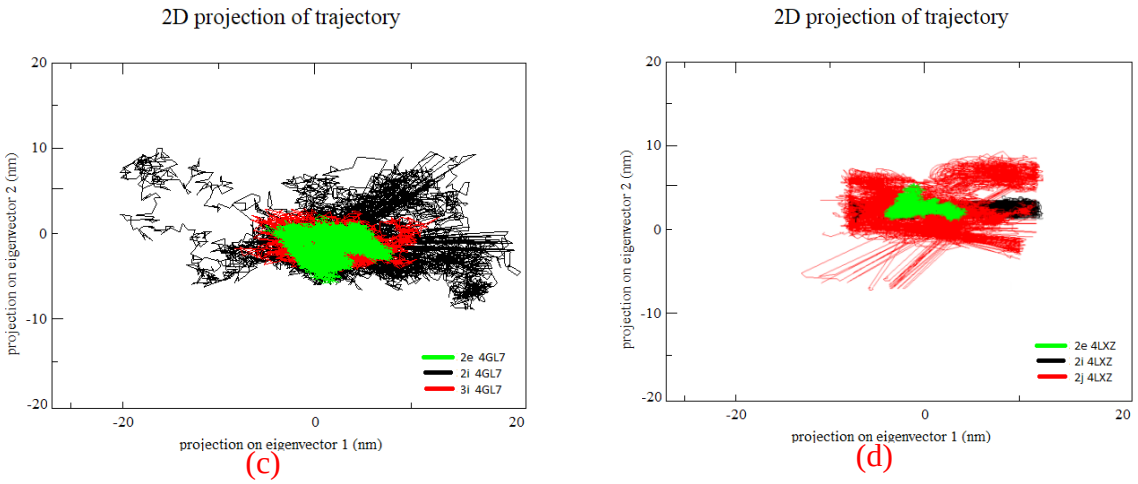
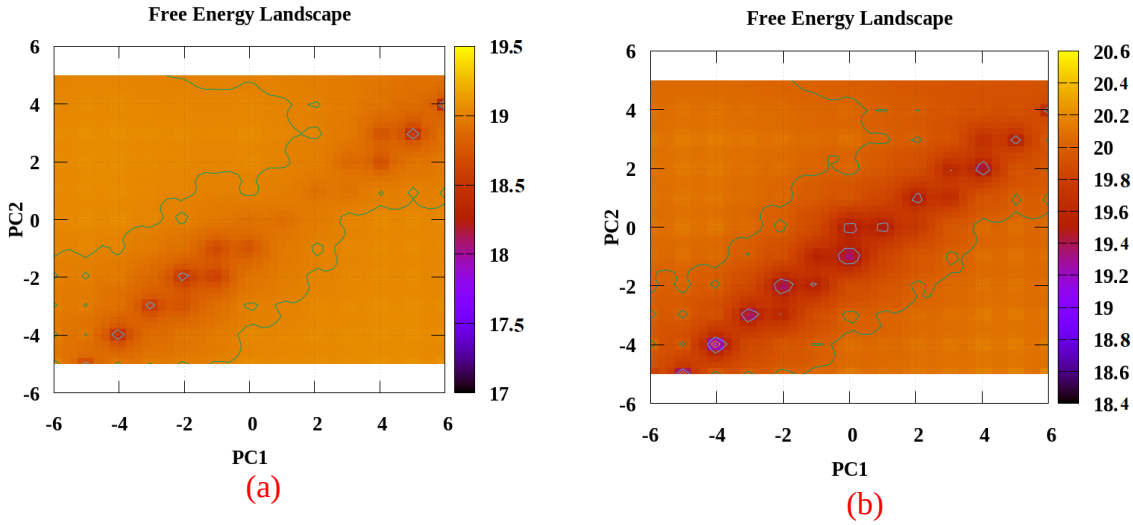
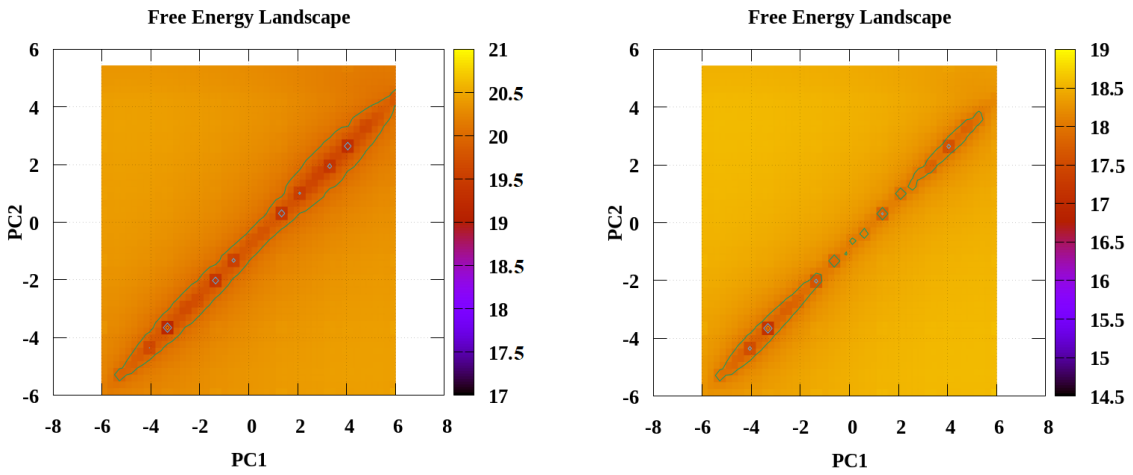


Figure 9: Scatter plot of principal component analysis (PCA) by projecting the eigenvectors of the stable complexes associated to the proteins: (a) 1M17, (b) 6GUE, (c) 4GL7, and (d) 4LXZ.

Subsequently, we generated the free energy surface or landscape (FES) for the selected complexes to unravel the differences in their protein folding patterns [26]. As depicted on Figure 10, the FES profile shows that all systems ultimately reached energy minima. However, the shifts in the positions of these minima indicate subtle changes in the conformation of the target proteins, which are attributed to the interaction with ligands.





(c) (d)

Figure 10: Free energy surface (FES) plot of the most stable complexes:
(a) 1M17-2h, (b) 4LXZ-2j, (c) 4GL7 -2j, and (d) 6GUE-3c.

One can note that free energy depicted on Figure 10, is computed high, however, as reported by previous related work [26,64,65], protein-ligand interactions occur on similar variable scale and depend strongly on the complexity of both the protein and ligand. On the opposite, more systems involving a sizeable transcriptional regulator (ZitR) complex [66], operate on a larger scale, influencing gene expression by binding to specific DNA sequences.

Solvent accessible surface area (SASA)

SASA is another parameter to analyze the nature of structural compactness and stability of a protein and its complex with the ligand through the course of MD simulation [41]. SASA of the twelve complexed proteins (1M17, 4LXZ, 4GL7, and 6GUE) combined with their ligands (2e, 2h, 2i, 2j, 3c, 3h, 3i) are shown on Figure 11.

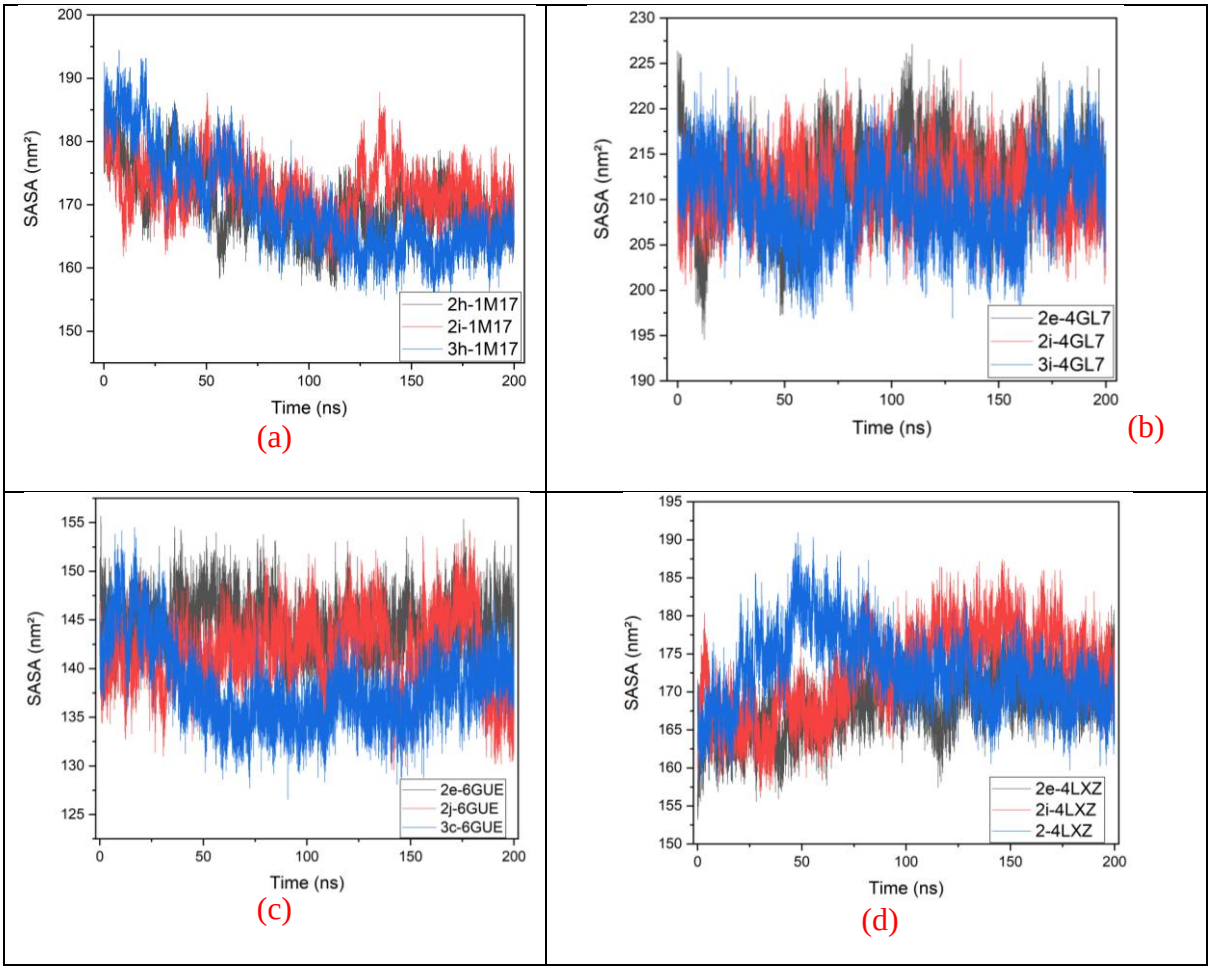


Figure 11: Solvent accessible surface area (SASA) of the proteins with ligands

All systems were found to be exhibiting a moderately constant SASA over the entire simulation period showing their stability. Indeed, in the case of the 2h-1M17, 2i-1M17, and 3h-1M17 complexes (Figure. 11a), the SASA average values are predicted equal to: 170, 172.4, and 169.8 nm², respectively, remaining remarkably constant throughout the timeline (0 – 200 ns) of simulation. These results indicate the structural stability of the 1M17 protein and its complexes under aqueous conditions, suggesting minimal compacting or expanding tendencies.

Similarly, for the 4GL7 protein, the 2e, 2i, and 3i complexes (figure. 10b), and the obtained SASA values are 212.43, 211.2, and 209 nm², respectively, exhibiting stable behavior throughout the simulation. In addition, the 2e-6GUE, 2j-6GUE, and 3c-6GUE complexes (figure. 10c), the SASA values i.e., 144.3, 141.7, and 138.07 nm², respectively, are predicted unchanged throughout the simulation. Finally, for the last series: 2j-4LXZ, 2e-4LXZ, and 2i-

4LXZ complexes (figure. 11d), the obtained values are 168, 172.8, 178.26, and 110.98 nm², respectively.

The SASA findings confirm the persistent stability of the ligand-receptor complexes within aqueous conditions, suggesting minimal alterations in structural compactness.

3.4. Prediction of new inhibitors

QSAR and molecular docking analysis

The structure of a compound determines its physicochemical properties as well as the ADMET (absorption, distribution, metabolism, excretion, and toxicity). The constructed MLR 2D-QSAR model was used to design twelve new hybrids (N1 – N12) of phenylsulfonyl furoxan and phenstatin derivatives based on the best docking score ligands (**2e**, **2i**, **2j**, **2h**, **3c**, **3h**, **3i**) associated with the four cell lines (4GL7, 6GUE, 1M17, 4XL7) (Table 5). The molecular structures of the twelve new ligands (N1 – N12) are modified according to the substitution R (in red) at the same site as shown on figure 12, and figure 13.

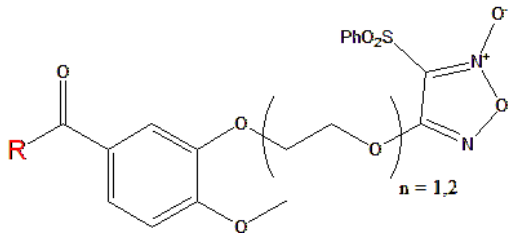


Figure 12: Structures of the twelve new N1 – N12 compounds

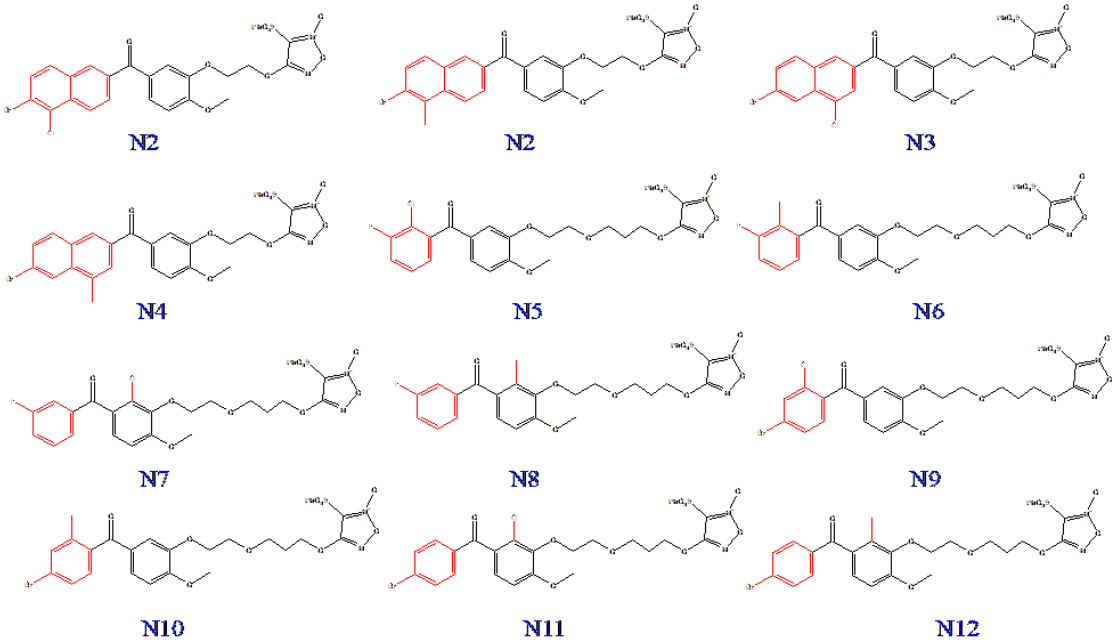


Figure 13: Structures of the twelve new N1 – N12 compounds.

The in silico results obtained from multiple linear regression (MLR) were also used to predict the anticancer activity (pIC50) of the twelve newly compounds against the four cell lines (A2780, MDA-MB-231, HCT-116, and A549), including their docking scores (kcal/mol) with their corresponding protein structures (4LXZ, 4GL7, 6GUE, 1M17).The result are gathered in Table 6.

Table 6: Newly designed (N1 – N12) compounds with predicted pIC50 values for the four cell lines (A2780, MDA-MB-231, HCT-116, and A549), and docking score (kcal/mol) with their protein structures (PDB codes: 4LXZ, 4GL7, 6GUE, 1M17).

Ligand	Predicted pIC50				Docking score (kcal/mol)			
	A2780	MDA-MB-231	HCT-116	A549	4GL7	6GUE	1M17	4LXZ
N1	6.24	4.88	4.97	5.55	-148.2	-144.7	-115.9	-145.5
N2	7.00	5.58	5.20	5.64	-148.1	-132.5	-130.2	-146.6
N3	6.23	4.82	5.07	5.51	-160.9	-139.5	-130.8	-136.4
N4	7.02	5.61	5.06	5.69	-143.2	-136.3	-116.2	-145.9
N5	5.78	4.24	4.71	4.79	-132.4	-120.7	-105.7	-101.5
N6	5.99	4.92	4.07	4.97	-135.1	-124.1	-123.2	-104.2
N7	5.65	4.26	4.19	4.84	-128.9	-121.6	-109.1	-104.1
N8	5.94	4.83	3.94	4.90	-125.5	-128.9	-101.9	-93.7
N9	5.90	4.52	4.79	5.03	-133.8	-128.5	-109.6	-110.5
N10	6.72	5.32	4.66	5.23	-138.6	-123.8	-106.7	-108.4
N11	5.90	4.49	4.87	5.01	-125.9	-122.3	-109.4	-101.9
N12	6.61	5.23	4.74	5.16	-132.4	-128.6	-107.0	-105.6

The results of Table 6 show that the predicted pIC50 values and docking score (kcal/mol) of the newly twelve designed models (N1 – N12), for the four cell lines i.e., A2780, MDA-MB-231, HCT-116, A549, are in good agreement with experimental in-vivo bioactivity activity (Table 1) reported for the actual hybrid derivatives [12]. It is noteworthy that among the in silico designed compounds, N2, N4, N10, and N12 ligands, exhibit the best-predicted pIC50 inhibition activity ranging from 4.66 to 7.02 values towards the four A2780, MDA-MB-231, HCT-116, A549 protein, and are associated with the 4-bromophenyl Me-derivative group (Table 6). Interestingly, their pIC50 values (N2, N4, N10, and N12), are closer to the best **3h**

ligand associated with the Ar = 4-bromophenyl group, which showed the most potent activities, ranging from 6.82 to 7.92 against both chemo-sensitive and resistant cancer cell lines, in particular towards the A2780 one (Table 1). More interestingly, the predicted inhibition activity of in silico-designed ligands (N1 – N12) is significantly higher than the observed prototype **1h** compound (Table 1). However, such predicted piC50 by the 2D-QSAR model, as expected to be found slightly lesser than those of in-vivo evaluated compounds [12].

Furthermore, the docking score shows that the binding energy values (kcal/mol) for predicted inhibitors (N1 – N12) are all negatively low, demonstrating a favorable conformation between ligands and the four proteins (4GL7, 6GUE, 1M17, 4LXZ). Notably, these new compounds (N1 – N12) have a significant affinity with the four enzyme-receptors producing H-bond interactions at the target point of amino acid residues. The binding energy ranging from –93 to –160 kcal/mol, reveals that four out of the twelve compounds, namely N1, N2, N3, and N4, show the best binding scores with the protein (4GL7, 6GUE, 1M17, 4LXZ) proteins.

Figure 14 shows the potent 2-D interactions between the best predicted docking scores (Table 6) between the (N1, N2, N3, and N4) ligands, and active site residues within the four N1-6GUE, N2-1M17, N3-4GL7, and N4-4LXZ complexes. The remaining 2-D interactions between the best predicted ligands N2, N3, and N4 with the 6GUE, 1M17, and 4LXZ proteins, are gathered on Figure S1 (See supplementary information).

It is noteworthy that the best-scoring N1, N2, N3, and N4 ligands, are predicted to function as competitive inhibitors of the considered enzymes, fitting into the substrate binding site and inactivating the enzyme’s catalytic activity.

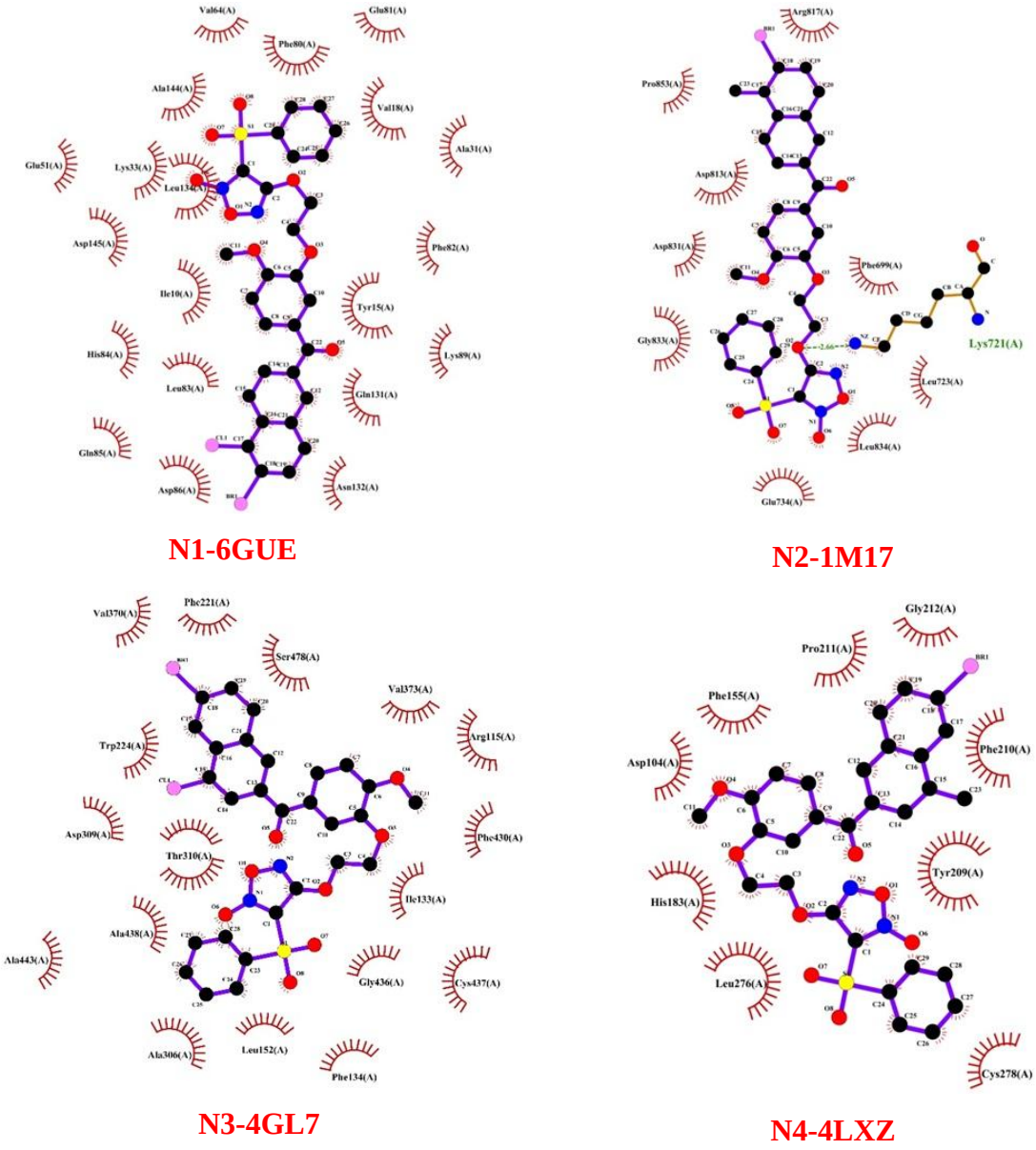


Figure 14**:** 2-D interactions between the active site residues of the four targets and selected ligands (N1-6GUE, N2-1M17, N3-4GL7, and N4-4LXZ)

Furthermore, figure 14 shows that the highest affinity of the designed N1, N2, N3, and N4 inhibitors towards the four (4GL7, 6GUE, 1M17, 4LXZ) targets, is due to the presence of van der Waals forces (e.g., Trp224, Thr310, Leu152, Leu477, Arg145, Arg115, and His183), which create a strong cohesive environment, thereby stabilizing the complexes formed. Furthermore, the molecular docking simulations with the four proteins show two strong H-bond interactions with Met374 and Arg115 amino acids.

In addition, steric interactions with LEU372-276-477, Phe134, Val370, Leu151, Ala306-443, Met303-446 and Thr310 residues, were seen and are almost identical to those found in the best docking scores for in-vitro compounds (table 5).

Regarding the docking scores and PCA findings (figure 14), the selected N1, N2, N3, and N4 ligands are predicted to be multitargeting inhibitors activity for Tyrosine kinase (1M17), Placental aromatase (4GL7), and Cyclin-dependent kinase (6GUE) targets, as obtained for the most stables 2e-4GL7 and 3i-4GL7 vs. 2e-6GUE and 2j-6GUE complexes, respectively (figure 9).

Indeed, in the case of multitargeting cancer therapy, inhibition activity for the aromatase enzyme has been found to be a viable targeted therapy for breast cancer [67]. Moreover, breast cancer tissues exhibit recurrent aromatase expression and generate elevated quantities of oestrogens in comparison to non-cancerous cells. It is noteworthy that enzyme, which belongs to the cytochrome P450 family, serves the purpose of catalyzing the final step of estrogen production [67].

Similarly, the inhibition of uncontrolled mitogen-activated protein kinase (MAPK) signaling pathway as well as PI3K-AKT (mTOR) downstream pathway, is related to different type of cancer therapy [68]. The inhibition of the protein ERK from the MAPK signaling pathway regulates the subcellular localization of CDK2 responsible of the human colorectal cancer [69]. Moreover, the epithelial growth factor receptor (EGRF) is associated with the aetiology of numerous human malignancies, such as lung's adenocarcinoma. Therefore, according to our result, the N3 molecule could exhibit potent EGFR inhibition activity for such type of cancer therapy [70]. One can note that EGFR inhibition in breast cancer, leads directly to the inhibition of ERK activity and reduces CDK2 activity [71]. In a similar way the targeted inhibition or reduction of CDK2/Cyclin A, and considering the Tyrosine kinase inhibitors (TKIs) as a targeted cancer therapy, the good docking score of N2-1M17 complex, would stops the cellular proliferative mechanism.

Therefore, except N4, the N1, N2 and N3 new designed ligands, could play an important role in the inhibition multitargeting (ovarian, breast, colorectal, and lung) cancer therapy.

3.5 Prediction of ADMET properties

The ADMET parameters (Absorption, Distribution, Metabolism, Excretion, and Toxicity) of the twelve new compounds were studied by the ADMET Predictor™ at specified dosage

level (10 mg/l), to predict several parameters, including metabolism, transporters, toxicity, and pharmacokinetic profiles for their possible genotoxicity based on the usual toxicology database. The ADMET results for new model compounds (N1 – N12) compared to the synthesized best in-vitro **3h** inhibition activity score ligand, are reported in table 7.

For targeting molecules oral administration, solubility is one foremost property that influences absorption to deliver an adequate quantity of active ingredients in a small volume. Moreover, many important physicochemical descriptors may correlate with pharmacokinetic (PK) or ADMET properties, such as water solubility (S) and lipophilicity or partition coefficient (Log P) which play major role in whether a drug can progress to be a successful drug candidate. Water solubility is given in log S (mol/L or mg/ml), and is referred to the following range: (insoluble ≤ 10 < poorly soluble ≤ 6 < moderately soluble ≤ 4 < soluble ≤ 2 < very soluble < 0 < very soluble) [26].

The ADMET results of the log P and log S values (Table 7) of almost designed (N1 – N12) compounds, indicate that they have a reasonable absorbency and are moderately water soluble, regarding the acceptable score of lipophilicity (–0.7 < LogP < 5), and solubility (0 < Log S < –6) as well as good polar surface areas (PSA ≤ 140) at specified dosage levels [26]. These results of the predicted bioavailability (Table 7) revealed the same scores for all newly designed twelve anticancer molecule candidates. It is noteworthy that for the intravenous route of administration, the bioavailability profile of the drug is nearly 100%.

Notably, compounds 6, 8, 10, and 12 have moderate water solubility (Log S = -3.0 mg/mL), and thus could facilitate good oral adsorption, relative to the less good scores for the other molecules. However, some ADMET and absorption risks are predicted by the ADMET_risk and absn_risk factors with a score in the range of 0-22 and 0-8, respectively, indicating the number of potential oral absorption problems, exceeding 7 (10%) and 4 (9%), respectively, a compound is likely to have.

Table 7**:** Pharmacokinetics prediction for the proposed (N1 – N12) compounds by ADMET Predictor™ cloud web server, and the best in vivo reference **3h** molecule.

Molecule		15j1Cl	15j1Me	15j2Cl	15j2Me	16c1Cl	16c1Me	16c2Cl	16c2Me	16h1Cl	16h1Me	16h2Cl	16h2Me
ID	**3h**	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12
MWt	620.461	660.913	640.495	660.913	640.495	593.995	573.577	593.995	573.577	654.906	634.488	654.906	634.488
LogP	2.080	3.184	3.184	3.184	3.184	2.080	2.080	2.080	2.080	2.276	2.276	2.276	2.276
Log S	-3.000	-5.514	-5.195	-5.684	-5.203	-5.311	-3.000	-5.349	-3.001	-5.500	-3.000	-5.567	-3.002
S+CL_Metab	Yes (45%)	Yes (57%)	Yes (55%)	Yes (57%)	Yes (55%)	Yes (49%)	Yes (45%)	Yes (46%)	No (44%)	Yes (55%)	Yes (45%)	Yes (55%)	Yes (45%)
PSA ($Å^2$)	138.27	129.04	129.04	129.04	129.04	138.27	138.27	138.27	138.27	138.27	138.27	138.27	138.27
Rule of 5	2	1	1	1	1	2	2	2	2	2	2	2	2
HBA	10	9	9	9	9	11	10	10	9	9	9	9	10
HBD	0	0	0	0	0	0	0	0	0	0	0	0	0
N_Rot	10	8	8	8	8	11	11	11	11	11	11	11	11
GI absorption	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
BBB_Filter	Low (79%) -1.309	Low (90%)	Low (90%)	Low (90%)	Low (90%)	Low (90%)	Low (74%)	Low (90%)	Low (74%)	Low (90%)	Low (84%)	Low (90%)	Low (79%)
ADMET_Risk	6.649	6.652	6.685	6.839	6.685	5.822	6.089	5.869	6.063	7.116	6.831	7.411	6.765
Absn_Risk	5,248	2.280	2.729	2.366	2.711	4.758	5.089	4.84	5.063	4.961	5.305	5.064	5.273
Pgp_Substr	Yes (97%)	Yes (84%)	Yes (88%)	Yes (84%)	Yes (86%)	Yes (93%)	Yes (97%)	Yes (95%)	Yes (97%)	Yes (97%)	Yes (99%)	Yes (97%)	Yes (99%)
Pgp_Inh	Yes (57%)	Yes (68%)	Yes (66%)	Yes (68%)	Yes (66%)	Yes (55%)	Yes (52%)	Yes (57%)	Yes (52%)	Yes (58%)	Yes (58%)	Yes (58%)	Yes (58%)

BCRP_Substr	No (62%)	No (47%)	No (47%)	No (60%)	No (60%)	Yes (95%)	Yes (84%)	Yes (95%)	Yes (84%)	No (58%)	No (67%)	No (58%)	No (67%)
OATP1B1_Substr	Yes (99%)	Yes (99%)	Yes (99%)	Yes (99%)	Yes (99%)	Yes (99%)	Yes (99%)	Yes (99%)	Yes (99%)	Yes (99%)	Yes (99%)	Yes (99%)	Yes (99%)
OATP1B1_Inh	Yes (72%)	Yes (84%)	Yes (84%)	Yes (84%)	Yes (84%)	Yes (79%)	Yes (72%)	Yes (79%)	Yes (72%)	Yes (84%)	Yes (76%)	Yes (84%)	Yes (72%)
CYP1A2_Inh	Yes (95%)	Yes (95%)	Yes (95%)	Yes (95%)	Yes (95%)	Yes (95%)	Yes (95%)	Yes (95%)	Yes (95%)	Yes (95%)	Yes (95%)	Yes (95%)	Yes (95%)
CYP1A2_Substr	No (85%)	No (70%)	No (70%)	No (63%)	No (72%)	No (61%)	No (65%)	No (61%)	No (59%)	No (78%)	No (85%)	No (78%)	No (81%)
CYP2A6_Substr	No (98%)	No	No (98%)	No (92%)	No (92%)	No (92%)	No (98%)	No (92%)	No (98%)	No (92%)	No (98%)	No (92%)	No (92%)
CYP2C8_Substr	Yes (52%)	Yes (60%)	Yes (60%)	Yes (60%)	Yes (60%)	No (49%)	No (49%)	Yes (49%)	No (46%)	Yes (56%)	Yes (54%)	Yes (56%)	Yes (54%)
CYP2C9_Inh	No (81%)	No (80%)	No (78%)	No (80%)	No (78%)	No (80%)	No (78%)	No (80%)	No (78%)	No (81%)	No (81%)	No (81%)	No (81%)
CYP2D6_Inh	No (80%)	No (70%)	No (65%)	No (70%)	No (65%)	No (70%)	No (67%)	No (70%)	No (65%)	No (80%)	No (76%)	No (80%)	No (76%)
CYP3A4_Inh	Yes (49%)	Yes (49%)	Yes (51%)	Yes (51%)	Yes (51%)	Yes (44%)	Yes (46%)	Yes (44%)	Yes (46%)	Yes (49%)	Yes (49%)	Yes (49%)	Yes (49%)
CYP3A4 Substr	Yes (98%)	Yes (98%)	Yes (98%)	Yes (98%)	Yes (98%)	Yes (98%)	Yes (92%)	Yes (98%)	Yes (92%)	Yes (98%)	Yes (98%)	Yes (98%)	Yes (98%)
CYP_Risk	0,171	1	1	1	1	0.054	0	0.029	0	0.642	0.317	0.662	0.25
TOX_Risk	1 (MUT)	2.372	1.956	2.473	1.981	1	1	1	1	1	1	1.149	1

Moreover, among physicochemical properties in Lipinski's rule [45], molecular weight (MW) slightly violated the threshold of 500, with a maximum number of H-bond acceptors of N10 HBA, except for the compound N5, and no H-bond donors (HBD) predicted. In addition, the number of rotatable bonds (n-Rot ≤ 10) is predicted slightly unfavorable for designed compounds N5 – N12.

Although two useful guideline rules of Lipinski's rule for drug-likeness are somewhat violated (MW and NO valence) for orally bioavailable designed compounds, the 'rule-of-five' has to some extent been overemphasized for designed compounds and even for synthesized moieties [72]. Furthermore, the in silico results for the designed compounds, correlate fairly well with that of the best in-vitro **3h** one (Table 7).

P-glycoprotein (P-gp) is responsible for efflux across biological membranes of a wide range of therapeutic drugs. One major role of P-gp is to protect the central nervous system (CNS) and cells from the harmful effects of drugs by transporting toxins and xenobiotics out of cells. It should be noted that all new compounds (N1 – N12) have a high probability (accuracy ca. 90%) of being a substrate of P-gp. Interestingly, the Pgp_Inh factor predicts whether or not the compound is a P-glycoprotein inhibitor (accuracy ca. 55%).

Moreover, the permeation of the blood-brain barrier (BBB) is a very important property in the pharmaceutical field because it determines whether or not a compound can cross the BBB and thus exert its therapeutic effect on the brain [73]. The standard value for the permeability of the BBB is good if its value is higher than 0.3 and bad if it is lower than −1 [74]. From the report on the BBB (table 7), it is clear that all the selected compounds have a moderate BBB permeability.

Also, according to the ADMET results and based on the drug similarity studies [75], the new model molecules (N1–N12) show acceptable human gastrointestinal (GI) absorption, metabolism properties, low total clearance, and small toxic properties, suggesting that these compounds are expected to exhibit good oral bioavailability, and behave as anti-cancer inhibitor drug candidates.

Notably, the clearance parameter (S+CL_Metab) corroborated the critical role of metabolism for the studied molecules. Indeed, their S+CL_Metab factor which predicts whether or not the clearance mechanism is metabolism, is predicted favorable with a range of 44 - 57% of accuracy.

It is noteworthy that Cytochrome P450 enzymes (CYPs) are a class of membrane-bound enzymes that contain heme. These enzymes are primarily found in the smooth endoplasmic reticulum and mitochondria of hepatocytes, as well as in the intestines. A total of 57 cytochrome P450 (CYP) isoforms have been identified in mammals, which play a crucial role in the oxidative metabolism of both xenobiotics and endogenous compounds [76]. Among these isoforms, five specific ones, namely CYPs 3A4, 2D6, 2C19, 2C9, and 1A2, are responsible for metabolizing over 80% of drugs commonly used in clinical practice [76]. Furthermore, the cytochrome P450 isoenzymes are important for drug metabolism in the liver [76].

Indeed, almost of newly designed compounds, were predicted to be substrates and inhibitors for several metabolizing enzymes of cytochrome P450 (CYP) i.e., CYP2C8, CYP3A4, for the breast cancer resistance protein (BCRP), and for the OATP1B1, OATP1B1 transponders, but fail to inhibit the human cytochromes: CYP2A6, CYP2C9, CYP2D6. Notably, in the family of CYP enzymes, CYP3A4 is an isoform of cytochrome P450, which is an enzyme responsible for the crucial detoxification of the human body and responsible for the modification of the pharmacokinetics of drugs. Moreover, CYP3A4 was the most important enzyme on account of metabolizing 50% of all drugs by itself [47].

Furthermore, it was predicted that overall newly designed compounds would likely exhibit inhibitory effects towards the CYP1A2 and CYP3A4 enzymes. Indeed, the analysis of the eight widely recognized CYP1A2 inhibitors with high affinity, shows that newly designed ligands (N1 – N12), possess common structural characteristics such as the presence of multiple aromatic moieties, heterocycles, secondary amines, and halogens. Moreover, high-affinity inhibitors of the CYP3A4 enzyme, that are commonly prescribed, exhibit a substantial molecular weight and possess a significant surface area, a considerable number of rotatable bonds, and hydrogen bond acceptors [76].

Interestingly, some studies show that most of the molecules fail in clinical trials due to their toxicity or poor pharmacokinetics [76]. Indeed, ADMET evaluation of CYP risks connected with P450 oxidation is predicted low (not exceeding 2.0), as a score in the 0-6 range indicating the number of potential problems a compound might have due to metabolism by one or more of major cytochrome P450s. In addition, the ADMET risk connected with toxicity is given by the TOX_Risk parameter and is also predicted low (not exceeding 2.0), as a score in the 0-6 range indicates the number of potential toxicity problems a compound

might have. Furthermore, the toxicity prediction, show that, in reference to the best in vitro score **3h** compound, the overall designed ligands (N1 – N12), which some of them are found to have mutagenicity (MUT), they exhibit low toxicity side (Table 7).

5. Conclusion

In this study, thirty-one hybrids of phenylsulfonyl furoxan and phenstatin derivatives (**1a-j**, **2a-j**, **3a-j**, **4** and **5**), were computationally analyzed for the first time using DFT calculations, and in silico approaches combining 2D-QSAR models, molecular docking and dynamics simulations, with support of ADMET properties for drug-likeness. Our study aims to develop in silico models able to design new multitargeting molecules that may be used as anti-cancer inhibitors for multiple cell lines. Indeed, based on the experimental pIC50 inhibition activity, the 2D-QSAR predictive model, combining the MLR technique and cross-validated by internal and external approaches, i.e., Y-randomization test and applicability domain. This approach was used satisfactorily to in silico evaluate the anticancer activity of considered hybrid compounds against the four different human tumor cell lines, namely ovarian cancer (A2780), breast cancer (MDA-MB 231), lung adenocarcinoma (A549), and colorectal cancer (HCT-116). Subsequently, the molecular docking analysis showed that the 2j-4GL7, 3c-6GUE, 3h-1M17, and 2j-LX4 protein complexes, exhibit the best binding energy of −174.681, −146.157, −126.561, and −140.314 kcal/mol, respectively. The R^2 values ranging from 0.74 to 0.85, reflects the good correlation between the predicted and observed (pIC50) activities. The dynamics simulations undertaken over a timeline of 200 ns, reveal satisfactorily RMSD, RMSF, Rg, SASA, PCA and FES values supporting their stability under physiological conditions. The developed predictive in silico-based model was used to design twelve novel hybrids (N1 – N12), and their potent anticancer activity was assessed. The molecular docking and dynamics simulations results for the best bonding affinity scoring (N1, N2, N3, and N4) ligands, correlate well with those obtained for the best selected experimental compounds (**2j**, **3c**, **3h**, and **2j**), which suggests their ability to inhibit the considered multitargeting proteins. Moreover, the ADMET prediction of the pharmacokinetic properties shows relatively reasonable oral bioavailability and low toxicity, but poor BBB permeability with respect to the reference ligand (**3h**).

These in silico findings could guide further research in the development and rational design of multitargeting anti-cancer drugs, in combination with in vitro and in vivo analysis,

providing promising pathways for the design of novel compounds with improved anticancer activity and reduced toxicity.

Acknowledgments

The authors are grateful to the Frères Mentouri University of Constantine 1 (Algeria) and the Pharmaceutical Sciences Research Center (CRSP) for providing computing facilities. Are also acknowledged, the General Directorate of Scientific Research and Technological Development (DGRSDT) for the PRFU project (2022-2024: Grant No. B00L01EN250120220001) and the Thematic Research Agency in Health and Life Sciences – ATRSSV for the PNR (Grant 2022-2024).

Statements and Declarations Funding:

The authors reported there is no funding associated with the work featured in this article.

Disclosure statement.

The authors have no competing interests to declare that are relevant to the content of this article.

Data availability statement.

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

Author contributions

Abdelmadjid Guendouzi and **Abdelkrim Guendouzi**: Conceptualization, Investigation, Methodology, Software, Material preparation, Data Collection, Writing – original draft; **Lotfi Belkhiri**: Formal analysis, Project administration, Supervision, Writing – original draft; **Farah Djelti**: analyzed and interpreted data, materials, analysis tools, and data; **Zineddine Mohamed Zendaoui**: analysis tools, and data; **Houari Brahim**: analyzed and interpreted data, tools, and data; **Abdelhamid Djekoun** and **Abdou Boucekkine**: Validation, Review & Editing.

All authors read and approved the final manuscript.

ORCID

Abdelmadjid Guendouzi http://orcid.org/0000-0001-7138-787X

Lotfi Belkhiri http://orcid.org/0000-0002-6093-6965

Abdelkrim Guendouzi http://orcid.org/0000-0001-9476-0489

Farah Djelti: https://orcid.org/my-orcid?orcid=0000-0002-8753-6443

Houari Brahim: https://orcid.org/0000-0002-1855-2706

Zineddine Mohamed Zendaoui https://orcid.org/0009-0007-1071-6271

Abdelhamid Djekoun http://orcid.org/0000-0001-5913-0801

Abdou Boucekkine https://orcid.org/0000-0002-3714-7191

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA: A Cancer Journal for Clinicians. 2023;73:17-48.
2. Huang J, Chan WC, Ngai CH, Lok V, Zhang L, Lucero-Prisno DE, Xu W, Zheng Z-J, Elcarte E, Withers M, Wong MCS, Universities oboNGHRGoAoPR. Worldwide Burden, Risk Factors, and Temporal Trends of Ovarian Cancer: A Global Study. Cancers. 2022;14:2230.
3. Anderson BO, Ilbawi AM, Fidarova E, Weiderpass E, Stevens L, Abdel-Wahab M, Mikkelsen B. The Global Breast Cancer Initiative: a strategic collaboration to strengthen health care for non-communicable diseases. The Lancet Oncology. 2021;22:578-81.
4. Varma DA, Singh M, Wakode S, Dinesh NE, Vinaik S, Asthana S, Tiwari M. Structure-based pharmacophore mapping and virtual screening of natural products to identify polypharmacological inhibitor against c-MET/EGFR/VEGFR-2. Journal of Biomolecular Structure and Dynamics. 2023;41:2956-70.
5. Stelitano G, Sammartino JC, Chiarelli LR. Multitargeting Compounds: A Promising Strategy to Overcome Multi-Drug Resistant Tuberculosis. Molecules [Internet]. 2020; 25.
6. Reddy AS, Zhang S. Polypharmacology: drug discovery for the future. Expert Review of Clinical Pharmacology. 2013;6:41-7.
7. Zięba A, Stępnicki P, Matosiuk D, Kaczor AA. What are the challenges with multi-targeted drug design for complex diseases? Expert Opinion on Drug Discovery. 2022;17:673-83.
8. Guendouzi A, Mekelleche SM. Prediction of the melting points of fatty acids from computed molecular descriptors: A quantitative structure–property relationship study. Chemistry and Physics of Lipids. 2012;165:1-6.
9. Nagamalla L, Kumar JVS. In silico screening of FDA approved drugs on AXL kinase and validation for breast cancer cell line. Journal of Biomolecular Structure and Dynamics. 2021;39:2056-70.
10. Chaudhari R, Tan Z, Huang B, Zhang S. Computational polypharmacology: a new paradigm for drug discovery. Expert Opinion on Drug Discovery. 2017;12:279-91.
11. Chaudhari R, Fong LW, Tan Z, Huang B, Zhang S. An up-to-date overview of computational polypharmacology in modern drug discovery. Expert Opinion on Drug Discovery. 2020;15:1025-44.

12. Huang X, Wang Y-S, Ma D, Wang Y-Y, Bian S-D, Zhang B, Qiao Y, He Z-R, Lv M, Cai G-L, Wang Z-X, Liu X-S, Shi J-B, Liu M-M. Synthesis and biological evaluation of novel hybrids of phenylsulfonyl furoxan and phenstatin derivatives as potent anti-tumor agents. European Journal of Medicinal Chemistry. 2022;230:114112.
13. Papa E, Dearden JC, Gramatica P. Linear QSAR regression models for the prediction of bioconcentration factors by physicochemical properties and structural theoretical molecular descriptors. Chemosphere. 2007;67:351-8.
14. Rücker C, Rücker G, Meringer M. y-Randomization and Its Variants in QSPR/QSAR. Journal of Chemical Information and Modeling. 2007;47:2345-57.
15. Thafar M, Raies AB, Albaradei S, Essack M, Bajic VB. Comparison Study of Computational Prediction Tools for Drug-Target Binding Affinities. Frontiers in Chemistry. 2019;7.
16. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. Journal of Cheminformatics. 2012;4:17.
17. Baerends EJ, Ziegler T, Autschbach J, Bashford D, Bérces A, Bickelhaupt F, Bo C, Boerrigter P, Cavallo L, Chong D. ADF2017, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands. ADF Available online: http://www scm com (accessed on 20 April 2020). 2014.
18. Becke AD. Density- functional thermochemistry. I. The effect of the exchange- only gradient correction. The Journal of Chemical Physics. 1992;96:2155-60.
19. Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Physical Review B. 1988;37:785-9.
20. Talete S. DRAGON for Windows and Linux. URL http://www talete mi it. 2007.
21. RStudio RT. Integrated development environment for R. RStudio, PBC: Boston, MA, USA. 2020.
22. Gramatica P. Principles of QSAR models validation: internal and external. QSAR & Combinatorial Science. 2007;26:694-701.
23. Golbraikh A, Tropsha A. Beware of q2! Journal of Molecular Graphics and Modelling. 2002;20:269-76.
24. Tropsha A. Best Practices for QSAR Model Development, Validation, and Exploitation. Molecular Informatics. 2010;29:476-88.
25. Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environmental Health Perspectives. 2003;111:1361-75.
26. Khamouli S, Belaidi S, Bakhouch M, Chtita S, Hashmi MA, Qais FA. QSAR modeling, molecular docking, ADMET prediction and molecular dynamics simulations of some 6-

arylquinazolin-4-amine derivatives as DYRK1A inhibitors. Journal of Molecular Structure. 2022;1258:132659.

27. Guendouzi A, Belkhiri L, Guendouzi A, Derouiche TMT, Djekoun A. A combined in silico approaches of 2D-QSAR, molecular docking, molecular dynamics and ADMET prediction of anti-cancer inhibitor activity for actinonin derivatives. Journal of Biomolecular Structure and Dynamics. 2023:1-15.
28. Cavasotto CN, Aucar MG. High-Throughput Docking Using Quantum Mechanical Scoring. Frontiers in Chemistry. 2020;8.
29. Schneider G. From Theory to Bench Experiment by Computer-assisted Drug Design. CHIMIA. 2012;66:120.
30. Viale M, Lentini G, Gangemi R, Castagnola P, Milani G, Ravera S, Bertola N, Carrieri A, Cavalluzzi MM. Lubeluzole Repositioning as Chemosensitizing Agent on Multidrug-Resistant Human Ovarian A2780/DX3 Cancer Cells. Molecules [Internet]. 2022; 27.
31. Dhumad AM, Jassem AM, Alharis RA, Almashal FA. Design, cytotoxic effects on breast cancer cell line (MDA-MB 231), and molecular docking of some maleimide-benzenesulfonamide derivatives. Journal of the Indian Chemical Society. 2021;98:100055.
32. Sulistyowaty MI, Widyowati R, Putra GS, Budiati T, Matsunami K. Synthesis, ADMET predictions, molecular docking studies, and in-vitro anticancer activity of some benzoxazines against A549 human lung cancer cells. 2021;32:385-92.
33. Ikwu FA, Isyaku Y, Obadawo BS, Lawal HA, Ajibowu SA. In silico design and molecular docking study of CDK2 inhibitors with potent cytotoxic activity against HCT116 colorectal cancer cell line. Journal of Genetic Engineering and Biotechnology. 2020;18:51.
34. Kouranov A, Xie L, de la Cruz J, Chen L, Westbrook J, Bourne PE, Berman HM. The RCSB PDB information portal for structural genomics. Nucleic Acids Research. 2006;34:D302-D5.
35. Bitencourt-Ferreira G, de Azevedo WF. Molegro Virtual Docker for Docking. In: de Azevedo Jr WF, editor. Docking Screens for Drug Discovery. New York, NY: Springer New York; 2019. p. 149-67.
36. Vanommeslaeghe K, MacKerell AD, Jr. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. Journal of Chemical Information and Modeling. 2012;52:3144-54.
37. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1:19-25.
38. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics. 1983;79:926-35.

39. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. The Journal of Chemical Physics. 2007;126.
40. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics. 1981;52:7182-90.
41. Ahmad S, Arsalan A, Hashmi A, Khan MA, Siddiqui WA, Younus H. A comparative study based on activity, conformation and computational analysis on the inhibition of human salivary aldehyde dehydrogenase by phthalate plasticizers: Implications in assessing the safety of packaged food items. Toxicology. 2021;462:152947.
42. Siddiqui S, Ameen F, Kausar T, Nayeem SM, Ur Rehman S, Tabish M. Biophysical insight into the binding mechanism of doxofylline to bovine serum albumin: An in vitro and in silico approach. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 2021;249:119296.
43. Ghose AK, Viswanadhan VN, Wendoloski JJ. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. Journal of Combinatorial Chemistry. 1999;1:55-68.
44. Egan WJ, Merz KM, Baldwin JJ. Prediction of Drug Absorption Using Multivariate Statistics. Journal of Medicinal Chemistry. 2000;43:3867-77.
45. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3–25.1. Advanced Drug Delivery Reviews. 2001;46:3-26.
46. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. Journal of Medicinal Chemistry. 2002;45:2615-23.
47. Okumu A, DiMaso M, Löbenberg R. Computer simulations using GastroPlus™ to justify a biowaiver for etoricoxib solid oral drug products. European Journal of Pharmaceutics and Biopharmaceutics. 2009;72:91-8.
48. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific Reports. 2017;7:42717.
49. Lengauer T, Rarey M. Computational methods for biomolecular docking. Current Opinion in Structural Biology. 1996;6:402-6.
50. Westermaier Y, Barril X, Scapozza L. Virtual screening: An in silico tool for interlacing the chemical universe with the proteome. Methods. 2015;71:44-57.

51. Laskowski RA, Swindells MB. LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. Journal of Chemical Information and Modeling. 2011;51:2778-86.
52. Chtita S, Belaidi S, Qais FA, Ouassaf M, AlMogren MM, Al-Zahrani AA, Bakhouch M, Belhassan A, Zaki H, Bouachrine M, Lakhlifi T. Unsymmetrical aromatic disulfides as SARS-CoV-2 Mpro inhibitors: Molecular docking, molecular dynamics, and ADME scoring investigations. Journal of King Saud University - Science. 2022;34:102226.
53. Divya V, Pushpa VL, Manoj KB. Cyclin dependent kinase 4 inhibitory activity of Thieno[2,3-d] pyrimidin-4-ylhydrazones – Multiple QSAR and docking studies. Journal of Molecular Structure. 2019;1183:263-73.
54. Abchir O, Daoui O, Belaidi S, Ouassaf M, Qais FA, ElKhattabi S, Belaaouad S, Chtita S. Design of novel benzimidazole derivatives as potential α-amylase inhibitors using QSAR, pharmacokinetics, molecular docking, and molecular dynamics simulation studies. Journal of Molecular Modeling. 2022;28:106.
55. Suresh PS, Kesarwani V, Kumari S, Shankar R, Sharma U. Bisbenzylisoquinolines from Cissampelos pareira L. as antimalarial agents: Molecular docking, pharmacokinetics analysis, and molecular dynamic simulation studies. Computational Biology and Chemistry. 2023;104:107826.
56. Nour H, Abdou A, Belaidi S, Jamal J, Elmakssoudi A, Dakir M, Chtita S. Discovery of promising cholinesterase inhibitors for Alzheimer's disease treatment through DFT, docking, and molecular dynamics studies of eugenol derivatives. Journal of the Chinese Chemical Society. 2022;69:1534-51.
57. Muthuvel SK, Elumalai E, K G, K H. Molecular docking and dynamics studies of 4-anilino quinazolines for epidermal growth factor receptor tyrosine kinase to find potent inhibitor. Journal of Receptors and Signal Transduction. 2018;38:475-83.
58. Yousaf MA, Anwer SA, Basheera S, Sivanandan S. Computational investigation of Moringa oleifera phytochemicals targeting EGFR: molecular docking, molecular dynamics simulation and density functional theory studies. Journal of Biomolecular Structure and Dynamics.1-23.
59. Abchir O, Daoui O, Nour H, Yamari I, Elkhattabi S, Errougui A, Chtita S. Exploration of Cannabis constituents as potential candidates against diabetes mellitus disease using molecular docking, dynamics simulations and ADMET investigations. Scientific African. 2023;21:e01745.
60. Maadwar S, Galla R. Cytotoxic oxindole derivatives: in vitro EGFR inhibition, pharmacophore modeling, 3D-QSAR and molecular dynamics studies. Journal of Receptors and Signal Transduction. 2019;39:460-9.

61. Garkusha NA, Anikeeva OP, Bayıl I, Taskin-Tok T, Safin DA. DFT, ADMET, molecular docking and molecular dynamics studies of pyridoxal. Journal of the Indian Chemical Society. 2023;100:100926.
62. Akash S, Abdelkrim G, Bayil I, Hosen ME, Mukerjee N, Shater AF, Saleh FM, Albadrani GM, Al-Ghadi MQ, Abdel-Daim MM, Tok TT. Antimalarial drug discovery against malaria parasites through haplopine modification: An advanced computational approach. Journal of Cellular and Molecular Medicine. 2023;27:3168-88.
63. Nour H, Hashmi MA, Belaidi S, Errougui A, El Kouali M, Talbi M, Chtita S. Design of Acetylcholinesterase Inhibitors as Promising Anti-Alzheimer's Agents Based on QSAR, Molecular Docking, and Molecular Dynamics Studies of Liquiritigenin Derivatives. ChemistrySelect. 2023;8:e202301466.
64. Xu S, Yuan H, Li L, Yang K, Zhao L. Computational screening of potential bromodomain-containing protein 2 inhibitors for blocking SARS-CoV-2 infection through pharmacophore modeling, molecular docking and molecular dynamics simulation. Arabian Journal of Chemistry. 2024;17:105365.
65. Kar B, Dehury B, Singh MK, Pati S, Bhattacharya D. Identification of phytocompounds as newer antiviral drugs against COVID-19 through molecular docking and simulation based study. Journal of Molecular Graphics and Modelling. 2022;114:108192.
66. He X, Ni D, Zhang H, Li X, Zhang J, Fu Q, Liu Y, Lu S. Zinc-mediated conformational preselection mechanism in the allosteric control of DNA binding to the zinc transcriptional regulator (ZitR). Scientific Reports. 2020;10:13276.
67. Chumsri S, Howes T, Bao T, Sabnis G, Brodie A. Aromatase, aromatase inhibitors, and breast cancer. The Journal of Steroid Biochemistry and Molecular Biology. 2011;125:13-22.
68. Stefani C, Miricescu D, Stanescu-Spinu I-I, Nica RI, Greabu M, Totan AR, Jinga M. Growth Factors, PI3K/AKT/mTOR and MAPK Signaling Pathways in Colorectal Cancer Pathogenesis: Where Are We Now? International Journal of Molecular Sciences [Internet]. 2021; 22.
69. Shi XN, Li H, Yao H, Liu X, Li L, Leung KS, Kung HF, Lin MCM. Adapalene inhibits the activity of cyclin-dependent kinase 2 in colorectal carcinoma. Mol Med Rep. 2015;12:6501-8.
70. Inamura K, Ninomiya H, Ishikawa Y, Matsubara O. Is the Epidermal Growth Factor Receptor Status in Lung Cancers Reflected in Clinicopathologic Features? Archives of Pathology & Laboratory Medicine. 2010;134:66-72.
71. Lents NH, Keenan SM, Bellone C, Baldassare JJ. Stimulation of the Raf/MEK/ERK Cascade Is Necessary and Sufficient for Activation and Thr-160 Phosphorylation of a Nuclear-targeted CDK2*. Journal of Biological Chemistry. 2002;277:47469-75.

72. DeGoey DA, Chen H-J, Cox PB, Wendt MD. Beyond the Rule of 5: Lessons Learned from AbbVie's Drugs and Compound Collection. Journal of Medicinal Chemistry. 2018;61:2636-51.
73. Clark DE. In silico prediction of blood–brain barrier permeation. Drug Discovery Today. 2003;8:927-33.
74. Zhu L, Zhao J, Zhang Y, Zhou W, Yin L, Wang Y, Fan Y, Chen Y, Liu H. ADME properties evaluation in drug discovery: in silico prediction of blood–brain partitioning. Molecular Diversity. 2018;22:979-90.
75. Han Y, Zhang J, Hu CQ, Zhang X, Ma B, Zhang P. In silico ADME and Toxicity Prediction of Ceftazidime and Its Impurities. Frontiers in Pharmacology. 2019;10.
76. Redlich G, Zanger UM, Riedmaier S, Bache N, Giessing ABM, Eisenacher M, Stephan C, Meyer HE, Jensen ON, Marcus K. Distinction between Human Cytochrome P450 (CYP) Isoforms and Identification of New Phosphorylation Sites by Mass Spectrometry. Journal of Proteome Research. 2008;7:4678-88.